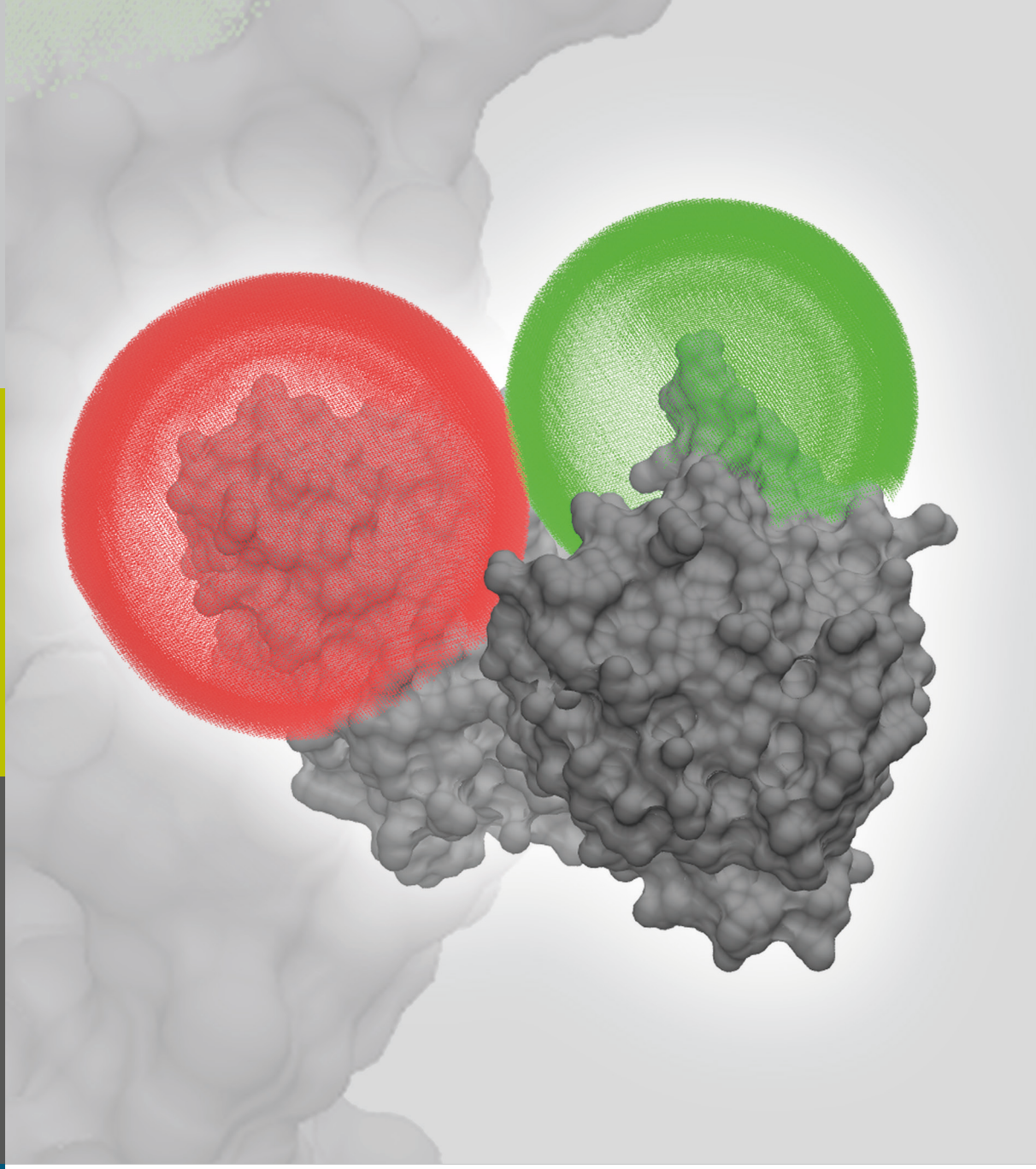


Interdomain Functional Dynamics of Phosphoglycerate Kinase Studied by Single-Molecule FRET

Matteo Gabba

Member of the Helmholtz Association

JÜLICH
FORSCHUNGSZENTRUM

Schriften des Forschungszentrums Jülich
Reihe Schlüsseltechnologien / Key Technologies Band / Volume 81

Forschungszentrum Jülich GmbH
Institute of Complex Systems
Molecular Biophysics (ICS-5)

Interdomain Functional Dynamics of Phosphoglycerate Kinase Studied by Single-Molecule FRET

Matteo Gabba

Schriften des Forschungszentrums Jülich
Reihe Schlüsseltechnologien / Key Technologies Band / Volume 81

ISSN 1866-1807 ISBN 978-3-89336-943-0

Bibliographic information published by the Deutsche Nationalbibliothek.
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at http://dnb.d-nb.de.

Publisher and Distributor:	Forschungszentrum Jülich GmbH Zentralbibliothek 52425 Jülich Tel: +49 2461 61-5368 Fax: +49 2461 61-6103 Email: zb-publikation@fz-juelich.de www.fz-juelich.de/zb
Cover Design:	Grafische Medien, Forschungszentrum Jülich GmbH
Printer:	Grafische Medien, Forschungszentrum Jülich GmbH
Copyright:	Forschungszentrum Jülich 2014

Schriften des Forschungszentrums Jülich
Reihe Schlüsseltechnologien / Key Technologies, Band / Volume 81

D 61 (Diss. Düsseldorf, Univ., 2013)

ISSN 1866-1807
ISBN 978-3-89336-943-0

The complete volume is freely available on the Internet on the Jülicher Open Access Server (JUWEL) at www.fz-juelich.de/zb/juwel

Ai miei genitori.

Veni, Vidi, Vici

Contents

Intents and Structure of the Thesis

The thesis reflects the steps followed to perform reliable single-molecule FRET (smFRET) experiments on phosphoglycerate kinase (PGK) from yeast. The main text is essential and divided in short chapters in order to relieve the reading. All technical details, demostrations, long descriptions, protocols, figures, and tables are reported in the appendix for the reader more interested in a deeper understanding. The first part encompasses Chap.3 to Chap.9 and is dedicated to the establishment of reliable single-molecule fluorescence measurements. Here, it is condensed the Ph.D. experience in a manual-like form for newcomers of the field willing to perform single-molecule experiments from scratch. Since high quality and controlled labeled samples are a fundamental prerequisite for single-molecule fluorescence measurements, this topic is discussed in Chap.3. The theoretical background for the comprehension of single-molecule detection in solution is given in a compact and consistent form in Chap.4. The basics about the three principal experimental approaches needed to perform single-molecule fluorescence measurements are presented in Chap.5, Chap.6 and Chap.7. The data treatment of the smFRET experiments is explained in Chap.8 and is mainly based on the works of Gopich and Szabo, and Seidel and co-workers. The procedure applied to determine the set of parameteres required for the analysis of the data is found in Chap.9. Here, a new strategy for the measurement of precise relative quantum yields on the confocal microscope is reported. Finally, in the last chapter (Chap.10) the whole knowledge is applied to study the interdomain dynamics of PGK. Here, the protein motions are modeled with an elastic network simulated under a random Multiparticle Collision Dynamics (rMPC) approach.

Chapter 1

Introduction

Living organisms are highly complex and organized systems able to replicate themself and to retain a state of nonequilibrium for a long period of time which is called life [1]. A large profusion of molecules with largely different dimensions, chemical nature and stabilities are continuosly moved, transformed and produced in a concerted flow carrying, processing and replicating a prodigious amount of information. In the middle of this tremendous chaos, very specific interactions and a complex network of coupled enzymatic reactions take place. At the heart of this network, energy is stored and consumed to sustain the chemical, osmotic and mechanical work necessary for life. D-glucose has a prominent position in the processes of energy transformation. In fact, D-glucose is the major energy supplier for most organisms, where during glycolysis the potential energy provided by this molecule is transferred to the highly energetic chemical bonds of ATP and NADH. Conversely, D-glucose is generated during gluconeogenesis. At the core of these two processes, the enzyme phosphoglycerate kinase (PGK) catalyzes the reversible phosphate transfer reaction between 1,3-bisphosphoglycerate (1,3-bPG) and ADP, with Mg^{2+} as a cofactor, generating ATP and 3-phosphoglycerate (3-PG) as products:

$$1,3\,\mathrm{bPG} + \mathrm{ADP} \overset{Mg^{2+}}{\rightleftharpoons} 3\,\mathrm{PG} + \mathrm{ATP} \tag{1.1}$$

Crystallographic studies show that PGK is a monomeric enzyme of $\sim 45\,\mathrm{kDa}$ with two compact and globular domains connected by a flexible hinge region (see Fig.1.1) [2]. This enzyme is highly conserved between different organisms and is required for ATP generation in the glycolytic pathways of aerobe and anaerobes, and for carbon fixation in plants [3]. The two domains, that fairly correspond to the N-terminal and C-terminal halves of the polypeptide chain, have similar size. Furthermore, they bind the substrates in two opposing regions situated at about 10 Å from each other, too far for the reaction to occur, therefore suggesting the possibility of a large scale hinge bending of the domains to bring the substrates together in

a water free environment for catalysis [3]. Namely, the nucleotide moieties (ATP and ADP) bind a hydrophobic pocket situated in the C-domain, and 3-PG or bBG bind a basic patch on the N-domain (Fig.1.1) [4]. The hinge-bending hypothesis is confirmed by the crystal structures of the open, closed, and partially closed conformations of PGK in the binary and ternary complexes with the substrates, that also brought about the idea of a substrate-induced conformational change [5]. In a step forward, small-angle X-ray scattering experiments support these hypothesis showing the closure of the cleft between the domains induced by the concerted interactions of ATP and 3-PG [6][7]. Conversely, binding of one ligand has small influence on the protein compactness.

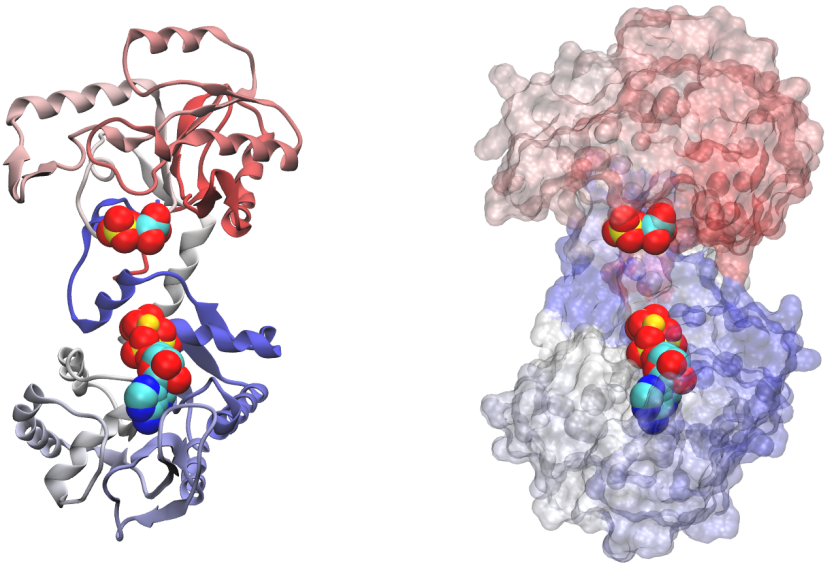


Figure 1.1: Phosphoglycerate kinase from yeast with the substrates, 3-PG and the ATP analogue AMP-PNP, respectively bound to the N- and the C-terminus (pdb:1qpg). The polypeptide chain is color coded from red (N-terminus) to blue (C-terminus).

Ensemble time-resolved fluorescence measurements give a more dynamic picture of PGK in solution [8]. In fact, they suggest the presence of a wide range of conformational substates visited by the protein domains interconverting on the slow nanoseconds timescale. Here, the ligands induce a change of the conformational equilibrium with a narrowing of the interdomain distance distributions. In addition, the ligands induce an increase of the mean interdomain distance which can be conceived by a hinge bending motion that involves the relative rotation of the two domains highlighting the importance of conformational flexibility in PGK. Molecular dynamics simulations of PGK in the apo form and in the ternary complex with ADP and bPG

also show hinge bending type motions on the nanoseconds time interval [9][10]. In the absence of the ligands the protein is more flexible, there are more hinge points that contribute to the hinge bending motions, and with the ligands it rigidifies, the hinge bending becomes strictly restrained with one dominant hinge point suggesting a higher directionality of the motions. Neutron spin echo spectroscopy (NSE) gives additional details about the intrinsic dynamics of PGK driven by thermal fluctuations of the interdomain configurations on the ~ 50 ns timescale [11]. The effect of substrates binding is observed in a small compaction of the partially closed average configuration, and in the reduction of the amplitude and timescale of the interdomain fluctuations. Furthermore, the character of the motions observed by the NSE experiments is described by a set of three normal modes: hinge-bending, rocking motion and propelled twist motion. The principal components analysis (PCA) of the ~ 100 ns trajectories obtained from MD simulations confirms the overall characteristics of these motions [12].

The explanation of enzymatic functional motions and protein-ligand interactions has evolved in the last decades from a mechanicistic static picture towards a statistical dynamic description. The evolution process has been driven by different theoretical approaches, new experimental techniques and the explosion of computational power. Especially, single-molecule techniques and NMR spectroscopies largely contributed by extending the range of observable length scales and timescales. Traditionally, under a crystallographic driven point of view, enzymes in solution are depicted as rigid units frozen in a well defined 3D structure. Here, the rigid body interconversion between static conformations, taking place around hinge regions and induced by the substrates, is described by the induced-fit model. The first step towards a dynamic description of the protein motions is a simple hydrodynamic model, that represents protein domains as mechanical entities accomplishing rigid body diffusive movements around flexible hinge regions [13]. The foundation of the modern view on protein dynamics is a statistical approach, the energy lanscape perspective [14]. Under this novel view, proteins in solution are considered as dynamic entities sampling a range of conformations [15][16] with collective large amplitude motions dictated by the topology of the interresidue contacts [17][18]. Here, the emerging paradigm is that *'structure-encodes-dynamics'*, and in turn *'dynamics-encodes-function'* [19]. Furthermore, the substrate binding mechanism is represented by a dynamic energy landscape [20], which supports the 'conformational selection' model. This model postulates that all the protein conformations pre-exist and the ligand selects the most favored conformers. Here, following binding, the system jumps from the ligand-free to the ligand-bound energy profiles, and the ensemble undergoes a population shift redestributing the conformational states [16]. In a step forward, large global structural tranformations between 'open' and 'closed' conformations of the domains are described combining the energy landscape viewpoint with the elastic model based on normal modes [21]. This theoretical framework puts forward a hierarchy of

timescales with fast local motions lubricating slower and large-scale functional movements of the domains [22]. Here, local flexible hot spots with fast dynamics on the picosecond to nanosecond timescale [22][23] and with frustrated interactions may act as hinges or may unfold to release elastic energy ('crack') and reassemble [24], facilitating the slower and large-scale motions [22].

A model system to study protein intrinsic motions and protein-ligand interactions is adenylate kinase (AK). In fact, this enzyme has been extensively analyzed with a wide range of experimental techniques and theoretical approaches [25][26][27][28][29][30]. The most comprehensive and detailed description is given in a nice and elegant paper by Henzler-Wildman et al. [23]. In this work, the authors clarified the dynamics of AK in the free and bound states over a large set of timescale (ns-ms) and length scales combining information from a set of complementery techniques (X-ray, NMR, MD, NMA, smFRET, NMR-EPR). Here, the free form of adenylate kinase samples a set of partially 'closed' conformations along the trajectory towards the catalytically competent 'closed' state in nanoseconds. Moreover, the fully closed conformation is rarely sampled on the microseconds-milliseconds timescale. Therefore, the larger-scale motions accomplished by free adenylate kinase are not random, but follow that pathway encoded in the fold that creates the configuration which better catalyzes the reaction. Finally, smFRET experiments on diffusing [23] and immobilized [27] AK highlights the role of the substrate binding. Here, binding reshapes the states distribution restricting the range of conformational fluctuations and shifting the equilibrium towards the catalytically relevant 'closed' conformation.

A global picture similar to AK can be inferred for PGK from the pieces of the puzzle obtained by independent techniques (X-ray, MD, NSE), as described previously in this section. Nevertheless, a characterization of possible large amplitude slow motions is still missing to get an overall description of the PGK catalytic behavior. In fact, for other enzymes these motions are the limiting step of the reaction [27]. In addition, a direct measurement of the substrate binding effects on the distribution of states is also missing. Since, the molecular weight of PGK is too high for NMR, single-molecule FRET based techniques are the methods of choice to answer these questions. In this thesis, smFRET experiments are performed on freely diffusing PGK labeled on the two different domains, in order to characterize the amplitude and the timescale of the interdomain motions. The substrate binding mechanism is also studied by looking at the redistribution of the conformational substates upon binding. From this data, a sketch of the PGK energy landscape is drawn, and a chemical kinetic model is proposed for the enzyme behavior in solution. In conclusion, the geometry of the interdomain motions and their relevance for catalysis is analyzed by an elastic network model simulated under the

random Multiparticle Collision Dynamics (rMPC) approach [31]. Here, the motions that are catalytically more relevant are identified and characterized.

Chapter 2

Materials

2.1 Chemicals

2.1.1 List of Chemicals

From Sigma-Aldrich Co., St.Louis, MI, USA.

- Sigmacote®
- CH_3COCH_3: acetone, CHROMASOLV® for HPLC $\geq 99.9\,\%$
- 3- PG: D-(-)-3-phosphoglyceric acid disodium salt, approx. $95\,\%$
- KADP: adenosine 5'-diphosphate monopotassium salt dihydrate, $\geq 95\,\%$ crystalline
- Mg_2ATP: adenosine 5'-triphosphate magnesium salt, $\geq 95\,\%$ bacterial
- $K_2HPO_4 \cdot 3\,H_2O$: potassium phosphate dibasic trihydrate
- DMSO: dimethyl sulfoxide, anhydrous, $\geq 99.9\,\%$
- GSH: L-glutathione reduced
- TCEP: tris(2-carboxyethyl)phosphine hydrochloride
- H_2SO_4: sulfuric acid, $(95.0 - 98.0)\,\%$
- GndHCl: guanidine hydrochloride, $\geq 99\,\%$
- H_2O_2: hydrogen peroxide solution, 35wt $\%$ in H_2O_2

From FLUKA Analytical, Sigma-Aldrich Chemie GmbH, Steinheim, Germany.

- $MgCl_2 \cdot 6\,H_2O$: magnesium chloride hexahydrate
- NaCl: sodium chloride

From RIEDEL-de Haën, Sigma-Aldrich Laborchemikalien GmbH, Seelze, Germany

- HCl: hydrochloric acid, min 32 %, purissi. p.a.

From AppliChem GmbH, Darmstadt, Germany.

- MOPS buffer grade
- TRIS buffer grade
- KH_2PO_4: potassium dihydrogen phosphate, pure Ph.Eu, NF
- $CHCl_3$: chloroform HPLC grade

From ROTH® GmbH, Karlsruhe, Germany.

- CH_3COCH_3: acetone, $\geq$ 99.5 % zur Synthese
- CH_3COCH_3: acetone, ROTIPURAN®, $\geq$ 99.9 % p.a. ACS ISO
- CH_3CH_2OH: ethanol, ROTISOLV® HPLC Gradient Grade
- NaOH: natriumhydroxid, $\geq$ 99 % p.a. ISO in Plätzchen
- $CaCO_3$: calcium carbonate, $\sim$ 99 % pulv.
- Tween®20

From MP Biomedicals, South Chillicothe Rd. Aurora, USA

- EDTA: ethylenediamine tetraacetic acid

2.1.2 Column Materials

From GE-Healthcare Bio-Sciences AB, Uppsala, Sweden.

- DEAE Sepharose, Fast Flow
- Pre-packed PD10 columns

From Sigma-Aldrich Co., St.Louis, MI, USA

- Sephadex®G-25

2.1.3 Fluorescent Dyes

From Molecular Probes® by Life Technologies™, USA.

- Al488mal: Alexa Fluor®488 C5-maleimide
- Al647mal: Alexa Fluor®647 C2-maleimide
- Al488COOH: Alexa Fluor®488 carboxylic acid succinimidyl ester, "mixed isomer"
- Al647COOH: Alexa Fluor®647 carboxylic acid succinimidyl ester
- Reference dye sampler kit:
  1. Quinine sulfate dihydrate in 0.1 M H_2SO_4
  2. Fluorescine in DMSO
  3. 5-carboxytetramethylrhodamine in DMSO
  4. Sulforhodamine 101 in DMSO
  5. Nile blue perchlorate salt in DMSO

From Atto-Tech GmbH, Siegen, Germany.

- Atto488NHS-ester
- Atto655NHS-ester

2.1.4 Proteins

The following proteins were genetically modified, expressed and purified in our molecular biology laboratory [32].

- PGK C97S Q135C: single-cysteine mutated phosphoglycerate kinase from yeast (*S. Cerevisiae*).
- PGK C97S S290C: single-cysteine mutated phosphoglycerate kinase from yeast (*S. Cerevisiae*).
- PGK C97S Q135C S290C: double-cysteine mutated phosphoglycerate kinase from yeast (*S. Cerevisiae*).

From Molecular Probes® by Life Technologies™, USA

- Strp-Al488: Streptavidin Alexa Fluor®488 conjugate
- Strp-Al647: Streptavidin Alexa Fluor®647 conjugate

2.1.5 DNA

From PURIMEX, Grebenstein, Germany.
Single stranded DNA fluorescentely labeled with Alexa dyes coupled to thymine (T = **X**) on position-5 via a C6- aminolinker.

- DNA modified at position 9 with T-C6-Alexa Fluor®488
  5'-ACT GAT CG**X** AAG CTA CTG AAG CGT A-3'
- DNA modified at position 7 with T-C6-Alexa Fluor®647
  5'-TAC GCT **X**CA GTA GCT TAC GAT CAG T-3'
- DNA modified at position 34 with T-C6-Alexa Fluor®488
  5'-GGA CTA GTC TAG GCG AAC GTT TAA GGC GAT CTC **X**GT TTA CAA CTC CGA-3'
- DNA modified at position 32 with T-C6-Alexa Fluor®647
  5'-TCG GAG TTG TAA ACA GAG ATC GCC TTA AAC G**X**T CGC CTA GAC TAG TCC-3'

From Eurofins MWG GmbH, Edersberg, Germany.
Unlabeled single stranded DNA complementary to the labeled one.

- 5'-ACT GAT CGT AAG CTA CTG AAG CGT A-3'
- 5'-TAC GCT TCA GTA GCT TAC GAT CAG T-3'
- 5'-GGA CTA GTC TAG GCG AAC GTT TAA GGC GAT CTC TGT TTA CAA CTC CGA-3'
- 5'-TCG GAG TTG TAA ACA GAG ATC GCC TTA AAC GTT CGC CTA GAC TAG TCC-3'

2.2 Buffers

Buffers were prepared by dissolving the chemicals in Milli-Q water. After adjusting the pH and the volume of the solution to the desired values, bacteria, fungi, and dust particles were filtered out using sterile filters. Only freshly prepared solutions were used for the measurements and the samples preparation.

- Labeling buffer
  100 mM KPi, 500 mM NaCl, pH 7.45

- PBS - Phosphate Buffer Saline
  50 mM KPi, 150 mM NaCl, pH 7.2

- Single-molecule FRET measurements buffer - smFRET buffer
  10 mM MOPS, 50 mM NaCl, 2 mM EDTA, pH 7.4

- DNA annealing and measuring buffer [33]
  20 mM TRIS, 100 mM NaCl, 10 mM $MgCl_2$, pH 7.5

- TE buffer
  10 mM TRIS, 1 mM EDTA, pH 7.5

2.3 Consumables

- 1.5 ml Protein-LoBind™ tubes, PCR clean (Eppendorf AG, Hamburg, Germany).

- 1.5 ml DNA-LoBind™ tubes, PCR clean (Eppendorf AG, Hamburg, Germany).

- 1.5 ml Safe-Lock™ tubes, PCR clean (Eppendorf AG, Hamburg, Germany).

- 1.5 ml Black-LiteSafe™ microcentrifuge tubes (Argos Technologies Inc., Elgin, IL, USA).

- Millex®-VV filter unit - Low protein binding - 33 mm Durapore® PVDF sterile membrane with 0.1 μm pore size (Merck Millipore Ltd., Tullagreen, Ireland).

- ACRODISC® 32 mm syringe filter with 0.2 μm Supor® membrane, Non-pyrogenic (Pall Corporation, Cornwall, UK).

- 13 mm filter device with NYLON filter medium and Polypropylene housing, 0.1 μm pore membrane (Whatman, GE-Helthcare Bio-Sciences AB, Uppsala, Sweden).

- Cover Slips, (20 x 20) mm #1.5 (Menzel-Gläser GmbH, Braunschweig, Germany).

- 2 absorption quarz cuvettes, $\ell = 50\,\text{mm}$, 3 ml, Q6 (Starna GmbH, Pfungstadt, Germany).

- 2 absorption quarz cuvettes, $\ell = 10\,\text{mm}$, 1 ml, Q6 (Hellma GmbH, Müllheim, Germany).

- 2 absorption quarz cuvettes, $\ell = 1\,\text{mm}$, 200 μl, QS (Hellma GmbH, Müllheim, Germany).

- 1 fluorescence quarz cuvette, $\ell = 3\,\text{mm}$, 45 μl, QS (Hellma GmbH, Müllheim, Germany).

2.4 Sample Storage

If not stated otherwise the samples were stored in normal 1.5 ml Safe-Lock tubes without shock freezing.

- Unlabeled PGK: -80 °C, shock freezing, 1.5 ml Protein-LowBind tubes, 10 mM MOPS buffer (pH 7.5)
- Labeled PGK: -80 °C, shock freezing, 1.5 ml Protein-LowBind tubes, 10 mM MOPS buffer (pH 7.5)
- Labeled Streptavidin: -20 °C, shock freezing, 200 μl PCR tubes, PBS
- Unlabeled DNA: -20 °C, TE buffer
- Labeled DNA: -20 °C, 1.5 ml DNA-LowBind tubes, DNA annealing buffer
- Atto655NHS and Atto488NHS: -20 °C, DMSO
- Al488NHS and Al647NHS: -20 °C, DMSO
- Al488mal and Al647mal: -80 °C, Black-LiteSafe™ tubes, lyophilized [34]

2.5 Slide Preparation Using Sigmacote®

To avoid protein sticking to the glass cover slides, the slides were treated with Sigmacote® prior to the measurements in order to create an hydrophobic and neutral thin layer [35]. The slides were immersed in ethanol and acetone HPLC grade for 10 min, respectively. After each step they were rinsed with Milli-Q water and dried with nitrogen. The cleaning procedure was followed by activation of the silanol group on the surface via plasma treatment for 3 min at 60 W. Finally, the Sigmacote® solution was deposited on top of the slides with a pipette, dried on a heating plate, and the excess was washed away with acetone and water. A test conducted with an FCS time series shows how the slides prepared in this way prevent protein sticking for at least one hour and even longer. In comparison with only plasma cleaned slides, the protein loss is reduced by $\sim 60\,\%$ immediately after pouring the sample on the sample holder ($t = 0$). The relative loss over one hour is reduced from $\sim 40\,\%$ to $\sim 12\,\%$.

2.6 List of the Instruments

- UV 2401 PC Spectrophotometer (Shimadzu, Duisburg, Germany).

- UV 2101 PC Spectrophotometer (Shimadzu, Duisburg, Germany).
- Spectrofluorophotometer QM-7 (Photon Technology International, Birmingham, NT, USA).
- pH-meter S20-SevenEasy™ mit InLab® Micro Electrode (Mettler Toledo, Giessen, Germany).
- "Zepto" plasma cleaner (Plasma Surface Technology, Diener Electronic GmbH, Ebhausen, Germany).
- J-810 Spectropolarimeter (JASCO GmbH, Groß-Umstadt, Germany).
- Abbe Refractometer AR3/AR4 with illumination unit (A. Krüss Optronic, Hamburg, Germany).
- AMVn Automated Micro Viscometer (Anton Paar GmbH, Graz, Austria)
- Megafuge 1.0R (Hereaus Sepatech, Thermo Scientific).
- PTC-200, Peltier Thermal Cycler (MJ Research, MA, USA).
- Microtime 200 (PicoQuant GmbH, Berlin, Germany)
    1. Microscope body: Olympus IX-81 (Olympus, Hamburg, Germany)
    2. Objective: UPLSAPO 60x, PlanApochromat, N.A. 1.2, W.D. 0.28, F.N. 26.5, water immersion, cover glass tickness 0.13-0-21, (400 - 900) nm (Olympus, Hamburg, Germany).
    3. Laser Driver: PDL828 "Sepia-II", Computer Controlled Multi-Channel Picosecond Pulsed Diode Laser Driver (PicoQuant GmbH, Berlin ,Germany).
    4. TCSPC-Acquisition unit: "HydraHarp-400", Picosecond Histogram Accumulating Real-Time Processor, Time-Correlated Single Photon Counting System with USB interface (PicoQuant GmbH, Berlin, Germany).
    5. Avalanche Photo Detectors - APD:
        (a) 2 x Photon Counting Module, SPCM-CD3307M and SPCM-AGR-14 (Perkin Elmer, Wood Bridge, Ontario, Canada).
        (b) 2 x Single Photon Counting Module, τ-SPAD, VLoK silicon avalanche photodiode (Laser Components, Olching, Germany) with quenching electronics (PicoQuant GmbH, Berlin, Germany).

6. Laser Heads (PicoQuant, Berlin, Germany):
   Red laser: LDH-D-C 640B
   Blue laser: LDH-D-C 470B
7. FCU-II Fiber coupling unit (PicoQuant, Berlin, Germany).
8. 50/50 Beam Splitter (Linos Photonics, Göttingen, Germany).
9. Polarizing Beam Splitter Cube (Linos Photonics, Göttingen, Germany).
10. Pinhole: 100 μm.
11. Major dichroic: 470/640 nm fitc/Cy5pc (AHF, Tübingen, Germany).
12. Minor dichroic: 600dcr (Chroma Technology, Bellows Falls, VT, USA).
13. Emission filters:
    Red: HQ 690/70M 202814 (Chroma Technology, Bellows Falls, VT, USA).
    Blue: XF 3003 520DF40 (Omega® Optical Inc., Battleboro, VT, USA)

2.7 List of the Software

The following software were used for data processing and analysis.

- MATLAB R2012a (The MathWorks Inc.)
- FMINUIT© for 64-bit Windows - A binding to Minuit for Matlab, Octave and Scilab [36]
- MMTK - The Molecular Modelling Toolkit [37]
- SymPhoTime® 5.3.2.2 (PicoQuant, Berlin, Germany)
- OriginPro v8.0988 (OriginaLab Corporation, Northampton, MA, USA)

Chapter 3

FRET Sample Preparation

3.1 Labeling of Phosphoglycerate Kinase

To perform single-molecule and ensemble FRET experiments, the elected protein should be labeled at the desired positions with functionalized fluorescent dyes. The dyes should specifically bind to the selected residues without perturbing the protein structure. This can be achieved by targeting the amino acid side chains exposed to the solvent. Two positions, at the distal end of the two domains of phosphoglycerate kinase (Fig.3.1), were chosen to study the protein interdomain dynamics, as already described in the literature [8][38].

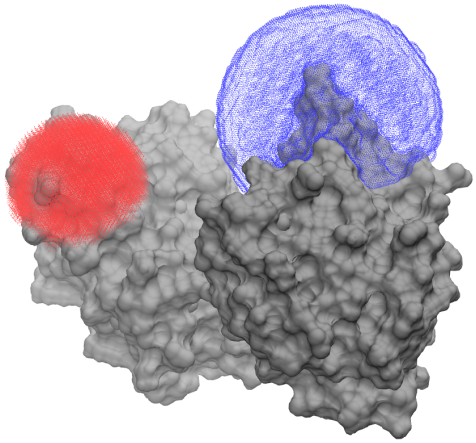


Figure 3.1: Crystal structure of phosphoglycerate kinase from yeast (pdb:1qpg) with dyes accessible volumes [33] calculated for C2-Al647mal (red) and C5-Al488mal (blue) attached at positions S290 and Q135, respectively. See Appx.E.7.3 for more detail about the accessible volume calculation.

Different chemistries and protocols can be used for the dye-protein coupling. For PGK the

general strategy described in [39] was followed. Residues at positions S290 and Q135 were mutated with cysteine after the substitution of the natural cysteine C97 by a serine. Single-cysteine mutants (smPGK), C97S S290C and C97S Q135C, as well as the double-cysteine mutant, C97S S290C Q135C (dmPGK), were prepared [32]. The mutants were labeled with C5-Alexa488 and C2-Alexa647 maleimide functionalized dye, as described in [32][35]. These two dyes represent a well known FRET pair. The double-labeled sample (dlPGK) was used for FRET-measurements and the single-labeled samples (slPGK) were employed to measure a host of parameters important for the data analysis, as described in Chap.9.
Here, the protocol described in [32] was slightly modified in order to have more control on the protein stability, to reduce aggregation, to avoid unspecific binding of the dye to the protein surface, to quantify more precisely the degree of labeling (DOL), and to better assess the sample quality.

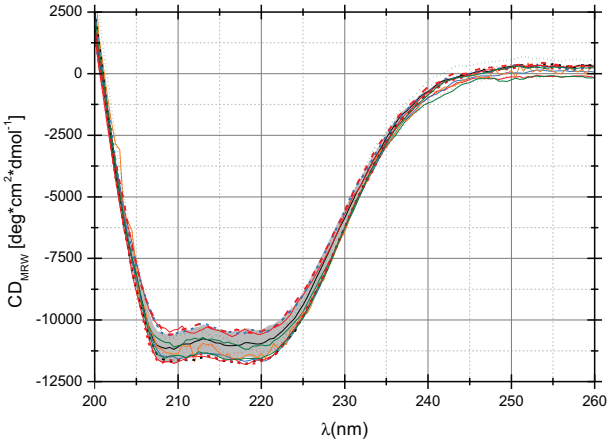


Figure 3.2: Measured CD-spectra of dmPGK and smPGK. The samples were stored in liquid nitrogen or at -80 °C. The grey shaded area represents the $\pm 5\,\%$ tolerance region with respect to the wtPGK spectrum shown in black. All samples were filtered with 0.1 μm-pore Whatman filters and the concentration was adjusted to an OD ~ 0.07. [Data provided by Tina Züchner]

3.1.1 Test of the Protein Stability

Quantitative and reproducible CD measurements [40][41] were systematically performed on the protein aliquots before starting the labeling reaction. If the quality of the sample was not satisfactory, new aliquots were checked and eventually a new stock was prepared. A deviation

of $\pm 5\,\%$ from the spectra of wild-type PGK (black solid line) was tolerated, as shown in Fig.3.2. Instability of samples stored for more than one year was observed.

3.1.2 Estimation of the Degree of Labeling

The degree of labeling (DOL) is defined as the ratio between the dye and the protein concentration measured with the absorption spectrometer:

$$\mathrm{DOL} = \frac{c_{dye}}{c_{prot}} \tag{3.1}$$

The degree of labeling is an important parameter to assess the quality of the labeling reaction and to improve the labeling conditions, because it tells how many dyes are bound to the protein. Even though, it must be complemented by FCS measurements to give reliable results (Sec.3.1.3). The DOL is related to the affinity ζ of the dye towards the labeling position. Assuming to have reproducible labeling conditions, the same dye will bind differently to separate targeted residues. Dye properties, as well as the features of the proteic environment surrounding the labeling site, will determine the fondness of the dye for the selected spot. Furthermore, diverse environments will differently affect the dye's physical behavior. For an example relative to PGK see Appx.D.4.3.

A series of physical quantities $\langle\xi\rangle$, which characterize the dye bound to dlPGK, must be measured in order to analyze the FRET measurements (see Chap.9). The value of $\langle\xi\rangle$ for one dye bound to dlPGK is the sum of the contribution of each labeling site ξ_i weighted by its relative population $\rho_i = \frac{r_i}{\sum r_i}$ which defines the occupation probability of the site itself:

$$\langle\xi\rangle = \sum \rho_i \cdot \xi_i \quad \text{(i=S290C,Q135C)} \tag{3.2}$$

Where, the population of each site r_i is proportional to the affinity ξ_i of the dye for the site itself: $r_i \propto \zeta_i$. Unfortunately, for the donor dye bound to the FRET-sample, it is not possible to directly measure $\langle\xi\rangle$ because of the energy transfer to the acceptor. Conversely, it is possible to measure ξ_i from the single-labeled samples. A way is then needed to estimate the relative population of each labeling site ρ_i in order to calculate $\langle\xi\rangle$ with Eq.3.2.

The relative populations ρ_i can be estimated from the measured values of the DOL_i under two assumptions. First, a linear relation is assumed between the DOL_i and the affinity ζ_i of each site: $\mathrm{DOL}_i \propto \zeta_i$. This assumption is reasonable if the dye concentration is not the limiting step of the labeling reaction, like for a reaction conducted with an excess of dye. Second, the affinity of the dye towards one site was considered to be the same on the single- and double-labeled samples: $\zeta_{i,\mathrm{slPGK}} \simeq \zeta_{i,\mathrm{dlPGK}}$. This second assumption is confirmed by the fact that $\mathrm{DOL}_{\mathrm{dlPGK}} \simeq \mathrm{DOL}_{\mathrm{S290C}} + \mathrm{DOL}_{\mathrm{Q135C}}$. For instance for Al647mal: $\mathrm{DOL}_{\mathrm{dlPGK}} = 0.63$

and $DOL_{S290C} + DOL_{Q135C} = 0.70$ (see Tab.3.1). Since, as stated before $r_i \propto \zeta_i$, the relative populations of a single dye bound to the double-mutant sample ρ_i are equal to the relative degrees of labeling of the two sites measured on the single-labeled samples:

$$\rho_i = \frac{r_i}{\sum r_i} = \frac{DOL_i}{\sum DOL_i} \tag{3.3}$$

This equation was applied to calculate the occupation probabilities (ρ_{S290C} and ρ_{Q135C}) of the two labeling sites of dlPGK for the two dyes (Al488mal and Al647mal) using the measured degrees of labeling (DOL_{S290C} and DOL_{Q135C}). The results are reported in Tab.3.1.

Table 3.1: Measured DOL_i and relative population ρ_i of slPGK. [Data provided by Daryan Kempe]

sample	dye	DOL_i	ρ_{Q135C}	ρ_{S290C}
Q135C	Al488mal	0.33	0.37	/
S290C	Al488mal	0.56	/	0.63
Q135C	Al647mal	0.24	0.34	/
S290C	Al647mal	0.46	/	0.66
$\langle\rho_i\rangle$			$\langle 0.36\rangle$	$\langle 0.64\rangle$

Two observations can be made. The relative affinity of the dyes towards each labeling site is independent from the dye properties: $\rho_{i,Al488} \simeq \rho_{i,Al647}$. The total affinity of Al488mal towards the labeling sites is greater than the affinity of Al647mal: $DOL_{Al488mal} \geq DOL_{Al647mal}$. This means that the relative distribution of the dyes between the labeling position is linked only to the features of the surrounding proteic environment. Conversely, the absolute binding probability depends on the physical-chemistry properties of the dyes. From this, the average $\langle\rho\rangle$ between $\rho_{i,Al488}$ and $\rho_{i,Al647}$ was calculated and used to estimate the desired physical properties by means of Eq.3.2:

$$\langle\xi\rangle = (0.36 \cdot \xi_{Q135C}) + (0.64 \cdot \xi_{S290C}) \tag{3.4}$$

The validity of Eq.3.3 and Eq.3.4 was tested by comparing the calculated and the measured amplitude averaged lifetimes of the acceptor in dlPGK: $\langle\tau_A\rangle_{x,calc} = 1.38\,ns$ and $\langle\tau_A\rangle_{x,exp} = 1.33\,ns$. The deviation is below $4\,\%$ (see Appx.D.4.4). The good agreement shows that the two assumptions, previously used to derive Eq.3.3, are correct.
Eq.3.4 is the key result of this section, because it allows one to calculate important physical parameters of the donor dye in the FRET-sample. These include the quantum-yield ϕ, the cross-talk α, the lifetimes τ and the Förster radius R_0, as described in Chap.9.

3.1.3 Test of the Labeled Sample with FCS

The results obtained with Eq.3.1 are misleading if not supported by diffusion coefficients properly measured by FCS (Sec.5.5). In fact, the DOL simply assesses the quantity of dye inside the measuring solution. Therefore, it provides no information about the specificity of the labeling reaction or about the presence of aggregates. Only a reproducible and stable diffusion coefficient gives this information. In fact, unwanted free dyes and aggregates are easily detected by an increase and a decrease of the diffusion coefficient with respect to the expected values.
Thereby, FCS was routinely measured on each sample, after the labeling procedure, and before starting every individual experiment. The hydrodynamic radius of wtPGK was measured with Dynamic Light Scattering ($R_h = 3.4\,nm$) at $\sim 22\,^{\circ}C$ in 10 mM MOPS buffer (pH 7.5) and the calculated (Sec.5.4) diffusion coefficient ($D_{wtPGK} = 68\,\frac{\mu m^2}{s}$) was used as a reference [Data provided by Daryan Kempe]. The average diffusion coefficient measured for double-labeled PGK in normal FRET buffer at circa $22\,^{\circ}C$ is $D_{dlPGK} = 55\,\frac{\mu m^2}{s}$. The difference with the calculated value is partially due to the temperature difference and partially to the presence of oligomers not resolved in the ensemble DLS measurement (Sec.10.5). Protein aggregates were generally observed at the end of the labeling procedure, therefore the aggregates were filtered out with a concentrator with 100 kDa cut-off. Furthermore, dye unspecific binding to partially unfolded PGK samples was also observed. Even unproperly stored maleimide functionalized dyes give rise to the same problem.

3.2 Labeled-DNA Hybridization Protocol

Double- and single-labeled double stranded DNA is needed to determine the detection efficiency ratio $g' = \frac{g_A}{g_D}$ of the confocal microscope, as described in Sec.9.7. Two samples with a distance of 10bp and 17bp between the labeling positions were used. Double stranded DNA, single- and double-labeled, was prepared by means of labeled and unlabeled single strands purchased from PURIMEX and Eurofins MWG GmbH, respectively. The desired ratio of the 0.1 mM stock solutions delivered by the companies were mixed in a total volume of $20\,\mu l$ DNA annealing buffer (20 mM TRIS, 100 mM NaCl, 2 mM EDTA, pH 7.4). The solution was then heated up to $98\,^{\circ}C$ for 3 min, and slowly cooled down to $25\,^{\circ}C$, with a temperature gradient of $0.1\,\frac{^{\circ}C}{s}$, using a PCR thermal cycler. To assemble double-labeled double-stranded DNA, the single-labeled single strands were mixed and hybridized with a 1 : 1 ratio. For the single-labeled samples, a 1.2 : 1 ratio was used, in order to have an excess of the unlabeled single strands.

3.3 Summary

A good labeled sample is an essential requirement to perform smFRET measurements. To assure that, two check points were added to the labeling protocol [32]. First, the protein structure was systematically controlled before each labeling reaction. Second, the samples were routinely tested with FCS to exclude dye unspecific binding and protein aggregation. An equation (Eq.3.4) based on two reasonable and verified assumptions was also derived to calculate a number of quantities (ϕ, α, τ and R_0) needed for the analysis of the FRET data. This equation will be extensively used in Chap.9.

Chapter 4

Single-Molecule Detection in Solution

4.1 The Golden Formulae in SMD

Abstracted from [42][43]

All physical quantities used in this thesis were attained with a fluorescence confocal microscope, on freeely diffusing molecules, with three basic experiments: FCS measurements (Chap.5), lifetime measurements (Chap.6), and single-molecule measurements (Chap.7). Here, the differences reside in the excitation pulsing scheme and in the configuration of the optical units of the detection system. Besides, the experiments can be additionally classified as ensemble ($\sim$ nM), and single-molecule measurements ($\sim$ pM). The main difference here is the average number N of molecules residing at the same time inside the detection volume. Despite the plethora of possible experimental conditions, a few equations are enough to grasp the principles of fluorescence detection.

These equations are derived under four hypotheses. First, the single-molecule detection limit (SMD) is examined. The probability to observe one molecule diffusing within the detection volume is significantly smaller than unity: $N << 1$. The extension to the ensemble case, with $N > 1$, is trivial. Second, the dye molecule is approximated as a system of three electronic energy levels (S_0: electronic ground state, S_1: first excited singlet-state and T_1: lowest excited triplet-state), as shown in Fig.4.2. Third, the space-dependent spatial detection efficiency Δ within the detection volume is almost constant ($\Delta \simeq$ const). Fourth, a laser excitation intensity I_e lower than the saturation intensity I_s is considered (Sec.4.1.1): $I_e \leq I_s$. Afterwards, this last assumption is relaxed in Sec.4.1.2 to introduce the contribution of the optical saturation.

4.1.1 Molecular Brightness: f_f

The molecular brightness f_f defines the dependence of the fluorescence detection rate per molecule from the properties of the diffusing dye and of the detection system. Consequently, it tells how to optimize the number of detected photons. With the hypothesis made in Sec.4.1, the fluorescence emission rate k_f $\left[\frac{\text{counts}}{\text{ms·molecule}}\right]$ of a single fluorophore excited by a laser intensity I_e is:

$$k_f = \phi_f \cdot \sigma_{01}(\lambda_{ex}) \cdot I_e \tag{4.1}$$

Where, the fluorescence quantum yield ϕ_f defines the probability of an absorbed photon to be reemitted radiatively (see Eq.9.4), and the molecular absorption cross-section $\sigma_{01}(\lambda_{ex})$ is the probability, for a photon with wavelength λ_{ex}, to excite the $S_0 \rightarrow S_1$ transition (see Fig.4.2). In Eq.4.1, the linear dependence $k_f \propto I_e$ holds only for $I_e \leq I_s$.
The molecular brightness f_f $\left[\frac{\text{counts}}{\text{ms·molecule}}\right]$, also called fluorescence detection rate, is derived from Eq.4.1 with a constant photon collection efficiency within the confocal volume ($\Delta \simeq \text{const}$), and taking into account the total fluorescence detection efficiency g_f:

$$f_f = g_f \cdot k_f \tag{4.2}$$

Where, $g_f = g_{opt} \cdot g_{el}$ is the fraction of emitted photons lost on the detection pathway. The term $g_{opt} = \int g\prime_{opt}(\lambda) \cdot f_{dye}(\lambda)\, d\lambda$ is the optical detection efficiency [44]. The optical detection efficiency g_{opt} is determined by the total transmission function of the optics $g\prime_{opt}(\lambda)$ (see Fig.4.1) and the fluorescence emission spectrum of the dye normalized with respect to the area $f_{dye}(\lambda)$. The contribution of the wavelength dependent detection efficiency of the APDs is included in g_{el}.
Finally, the number of detected fluorescence photons F_f $\left[\frac{\text{counts}}{\text{molecule}}\right]$, from a single molecule diffusing inside the confocal volume, can be calculated from the molecular brightness f_f integrated over the observation time T [ms]:

$$F_f = g_f \cdot k_f \cdot T \tag{4.3}$$

Eq.4.3 clearly displays how to tune the number of detected photons F_f by adjusting one of the three terms: g_f, k_f and T. The optical detection efficiency g_{opt} of the employed set-up is shown in Fig.4.1. The filters set is optimal for the Al488-Al647 FRET-pair and 470 nm / 640 nm laser excitation. A 100 μm pin-hole was used to have an observation times T of a few milliseconds for diffusing PGK. Finally, the excitation intensity I_e was optimized as described in Sec.4.1.2.
Up to now, an ideal detection system was considered. In real systems, solvent molecules, fluorescence impurities and APDs electronic noise are source of background. Theoretical rela-

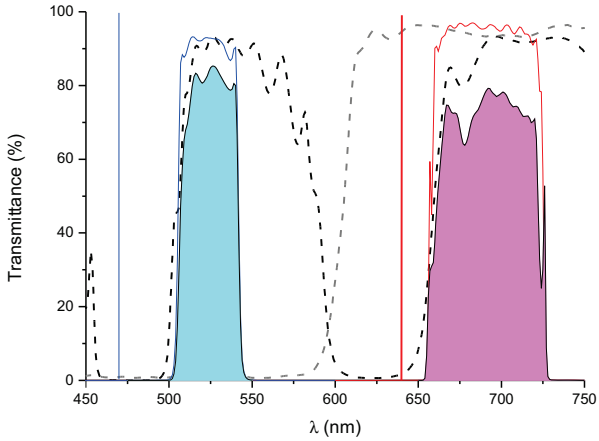


Figure 4.1: The shaded areas display the total transmission function of the optics $g'_{opt}(\lambda) = \prod T_i(\lambda)$ calculated from the measured transmission spectra $T_i(\lambda)$ of all the optical units of the Microtime 200. Major dichroic: 470/640 nm fitc/Cy5pc (black dashed line). Minor dichroic: 600dcr (gray dashes line). Blue emission filter: XF 3003 520DF40 (blue line). Red emission filter: HQ 690/70M (red line). The vertical lines show the laser excitation wavelengths: 470 nm and 640 nm.

tions can be derived in analogy with Eq.4.2 and Eq.4.3 [42]. For simplicity, the experimentally determined background detection rate bg_f $\left[\frac{counts}{ms}\right]$ and the number of background photons $BG_f = bg_f \cdot T$ $\left[\frac{counts}{molecule}\right]$ are considered here. These two contributions can be measured from the buffer solutions. Since in real experiments the background (bg_f and BG_f) and the fluorescence emitted (f_f and F_f) photons add up, the total detection rate s_f and the total number of detected photons S_f are calculated by adding the two terms:

$$s_f = f_f + bg_f \tag{4.4}$$

$$S_f = F_f + BG_f \tag{4.5}$$

These equations allow one to calculate the dye contributions, f_f and F_f, to the total detected intensity signal. The extension to the ensemble level ($\sim$ nM), with an arbitrary number N of not interacting molecules diffusing inside the confocal volume, is trivial. Eq.4.2 must simply be multiplied by N, and Eq.4.4 and Eq.4.5 consequently modified as follows:

$$s_f = N \cdot f_f + bg_f \tag{4.6}$$

$$S_f = N \cdot F_f + BG_f \tag{4.7}$$

The molecular brightness f_f can then be obtained from the experimentally determined s_f, bg_f and N (see Sec.5.3):

$$f_f = \frac{s_f - bg_f}{N} \tag{4.8}$$

4.1.2 Signal-to-background Ratio: SB_0

In the previous section, an assumption was made about the laser excitation intensity I_e, namely, I_e was considered to be lower than the saturation intensity ($I_e < I_s$). This assumption can be experimentally fulfilled only if the saturation intensity I_s is known. To determine the value of I_s, a power series must be measured, and the signal-to-background ratio SB_0 plotted against the laser power I_e. In this section, the condition $I_e < I_s$ is relaxed, and the empirical definition of SB_0 is given. Eq.4.1 must be rewritten to include the effect of the optical saturation:

$$k_f = \frac{\phi_f \cdot \sigma_{01}(\lambda_{ex}) \cdot I_e}{1 + I_e \cdot \tau \cdot \sigma_{01}(\lambda_{ex}) \cdot \left[1 + \frac{k_{isc}}{k_{ph}}\right]} \tag{4.9}$$

The inter-system crossing rate k_{isc} describes the population rate of the triplet-state T_1 due to the $S_1 \rightarrow T_1$ electronic transition (see Fig.4.2). The phosphorescence rate k_{ph} defines the depletion of triplet state T_1 towards the ground state S_0. Finally, the fluorescence lifetime $\tau \simeq k_{10}^{-1}$ [ns] indicates the average time spent by an electron in the excited state S_1 before its radiative decay to S_0 with a photon emission.

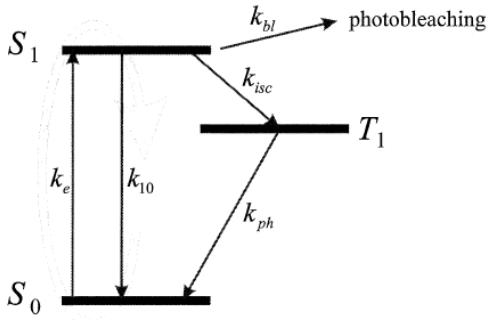


Figure 4.2: Jablonski diagram showing the main photo-physical transitions of a typical fluorescing molecule. [42]

The second term in the denominator of Eq.4.9 induces a saturation of the $S_0 \rightarrow S_1$ electronic transition at high excitation intensity, breaking down the linear relation between k_f and I_e. With this in mind, the signal-to-background ratio SB_0 is defined as the ratio between the

fluorescence detection rate f_f and the background detection rate bg_f:

$$SB_0 = \frac{f_f}{bg_f} \tag{4.10}$$

Eq.4.10 can now be used to determine the optimal excitation intensity I_{opt} before saturation occurs (see Sec.4.2).

4.2 Optimal Excitation Intensity

The saturation intensity I_s [a.u.] of a fluorescing molecule defines the maximum reachable laser excitation intensity before saturation and photo-destruction start to occur. I_s also sets the limit to the linear regime between the excitation intensity I_e and the molecular brightness f_f (Eq.4.2). For those reasons, I_s is a compromise between the needs of large numbers of detected photons and the fluorophore photo-protection. In general, the boundary between the linear and the saturation regime cannot be clearly identified. Therefore, an optimal excitation intensity I_{opt} is defined in order to be on the safe side.

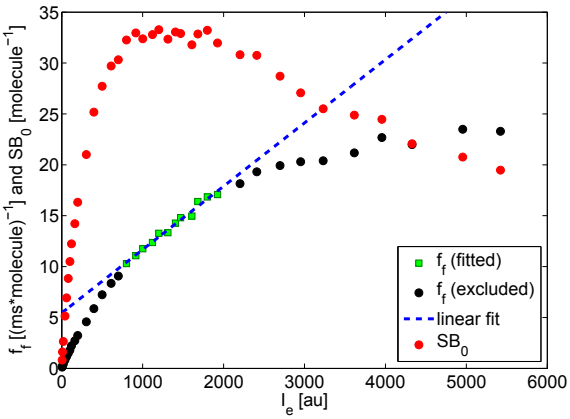


Figure 4.3: Representative plot of the molecular brightness f_f (black) and the signal-to-background ratio SB_0 (red) plotted against the laser excitation intensity I_e. The fitted linear region, which corresponds to the plateau, is shown in green. Data taken from A-slPGK S290C (Appx.A.1).

For each sample used in this thesis, the optimal excitation intensity I_{opt} was determined from the power series measured at ensemble level (Appx.A.1). The total count rate s_f and the background count rate bg_f were measured as a function of the laser excitation intensity I_e, and N was determined with an FCS measurement (see Sec.5.3). Thence, the molecular brightness f_f and the signal-to-background ratio SB_0 were calculated with Eq.4.8 and Eq.4.10, and plotted

against the laser intensity, as shown in Fig.4.3. As stated in [42] and displayed in Fig.4.3, the linear dependence between f_f and I_e corresponds to a plateau in the (SB_0, I_e) curve, which is shown in red. The (f_f, I_e) data were then fitted with a straight line, starting from the beginning of the plateau till the end of the plateau (green squares). As shown in Fig.4.3, the point of the fit at highest laser intensity corresponds to the saturation intensity I_s, 1900 a.u. in this case. The resulting values for all the samples used in this work are shown in Tab.4.1.

Table 4.1: Measured saturation intensities I_s for single-labeled DNA, PGK, Strp and free Atto-dyes. 10bp and 17bp refer to the base pair distance between the two dyes in the corresponding dlDNA FRET samples. A=Alexa647 and D=Alexa488. The full set of plots is displayed in Appx.A.1.

sample	I_s [a.u.]	sample	I_s [a.u.]	sample	I_s [a.u.]
A-slDNA 10bp	2400	A-slPGK Q135C	1900	A-Strp	2000
A-slDNA 17bp	2400	A-slPGK S290C	1900	D-Strp	700
D-slDNA 10bp	1400	D-slPGK Q135C	500	Atto655-NHS	2700
D-slDNA 17bp	1400	D-slPGK S290C	600	Atto488-NHS	2150

The values of Tab.4.1 correspond to the optimal excitation intensities I_{opt} for the single-labeled samples described by a three state energy diagram (Fig.4.2). When donor and acceptor are bound to the same molecule, the dipolar coupling open a deexcitation/excitation path between the excited states of the dyes: $S_{1,D} \rightarrow S_{1,A}$ (see Fig.6.1). Thus, the donor can stand a higher excitation intensity and the acceptor a lower one before saturation occurs. The energy fraction, relieved from the donor and charged on the acceptor shoulders, is defined by the energy transfer efficiency E (see Sec.8.2). This quantity defines the probability that the excitated donor state transfers its energy to the acceptor rather than decaying to its ground state. The optimal laser excitation intensities, I_D and I_A, can then be calculated under FRET coupling and PIE excitation (Sec.7.2) for the donor and the acceptor, respectively. Two conditions must be fullfilled at the same time:

$$(1 - E) \cdot I_D \leq I_{s,D} \tag{4.11}$$

$$E \cdot I_D + I_A \leq I_{s,A} \tag{4.12}$$

Where, $I_{s,D}$ and $I_{s,A}$ are the saturation intensities measured with the single-labeled samples, previously reported in Tab.4.1. For dlPGK, the donor laser excitation intensity I_D was chosen to be equal to $I_{s,D}$ and for dlDNA a little bit higher than $I_{s,D}$. The acceptor laser excitation intensity I_A was settled at the beginning of the plateau, as observed in the (SB_0, I_e) plots (Fig.4.3). The conditions were then verified using an approximated E value attained from test measurements. The intensities shown in Tab.4.2 were used to perform smFRET experiments.

Table 4.2: Optimal laser intensities used to perform the smFRET experiments under PIE excitation of the donor and the acceptor.

sample	I_D [a.u.]	I_A [a.u.]
dIDNA	2000	1000
dIPGK	1375	500

4.3 a.u. to Watt Conversion

In Sec.4.2, the laser intensity values are given in arbitrary units [a.u.] for the used experimental set up. The conversion functions of the laser power, from arbitrary units to Watt, were measured with a powermeter positioned at the back focal plane of the microscope's objective. The results are reported in Fig.4.4.

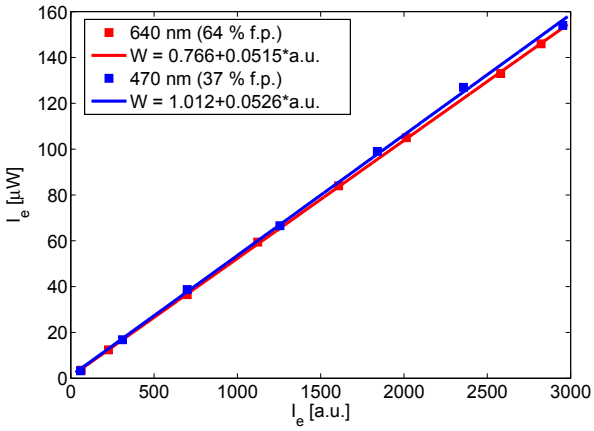


Figure 4.4: Measured conversion functions from a.u. to Watt laser intensities. The measurements were performed with the red (640 nm) and the blue (470 nm) lasers operating at 64 % and 37 % of full power (f.p.), respectively. The linear conversion functions from a.u. to Watts [μW] are reported in the figure inset.

4.4 Summary

The equations derived in this chapter are the foundation of most of the experiments described in this thesis. Eq.4.2 allows one to understand how to measure the quantum yield ϕ, the cross-talk α and the detection efficiency ratio $g' = \frac{g_A}{g_D}$. Eq.4.3 clearly shows how to optimize the number of detected photons in single-molecule experiments. Eq.4.8 permits one to meausure the molecular brightness from a FCS measurement, and to fix the threshold applied to the

inter-photon distance trace for single-molecule discrimination (see Sec.7.3). Finally, Eq.4.9 set a limit to the laser excitation intensity applicable to freely diffusing dyes. The optimal laser excitation intensities for the single-labeled and double-labeled DNA and PGK samples were measured by employing Eq.4.10, with the conditions set by Eq.4.11 and Eq.4.12 under FRET coupling (see Sec.4.2). Furthermore, the knowledge of the saturation intensity I_s is important for the experimentalist because it allows one to use simplified equations for data analysis, and prevents undesired photophysical effects.

Chapter 5

FCS Measurements

5.1 Introduction

Abstracted from [42][45][46].
Fluorescence correlation spectroscopy (FCS) measures the autocorrelation function of fluorescence intensity fluctuations in diluted samples ($\sim$ nM). Barring a few exceptions, almost every microscopic phenomenon able to induce fluorescence intensity fluctuations can be detected in a time range from a few picoseconds up to tens of milliseconds, and its characteristic correlation time τ_c can be measured. The limits are the timing accuracy of the detectors [47] and the maximum observation time. In standard FCS measurements, performed on molecules diffusing in aqueouse solution with a confocal microscope and pulsed excitation, the time interval is restricted to the μs $-$ ms time domain. In this time regime, the main contributions to the intensity fluctuations are the translational diffusion and the triplet-state kinetics. Disregarding the dye photo-physics, the main parameters that can be extracted from such a measurement are the average number of molecules within the observation volume N, the characteristic diffusion time τ_{DT}, also called dwell time, and the ellipticity κ of the observation volume. Those parameters depend on the experimental set-up (τ_{DT} and κ) and on the sample properties (N and τ_{DT}), like the concentration c and the diffusion coefficient D. Besides, the molecular brightness f_f (Sec.4.1.1), the radius of gyration R_g and the hydrodynamic radius R_h can be derived from a FCS measurement. Fluorescence correlation spectroscopy is thus a valuable tool, to assess the sample quality (Sec.3.1.3), to characterize the dye properties (Appx.D.4.2), and to determine large conformational changes in proteins (Sec.10.5).

5.2 FCS Principles

In macroscopic systems at equilibrium, the average physical parameters $\langle \xi(t) \rangle$ are constant in time. Conversely, if the concentration of molecules is not too high ($\sim$ nM), the physical quantities $\xi(t)$ do not average out, and fluctuations $\delta\xi(t)$ are observed. Therefore, the microscopic origin of the processes prompting the fluctuations $\delta\xi(t)$ can be characterized by correlation analysis.
In FCS, the fluorescence intensity $F(t)$ is measured as a function of time with a confocal microscope and highly sensitive avalanche photodiodes. The autocorrelation function $G(\tau)$ of the fluorescence intensity fluctuations $\delta F(t) = F(t) - \langle F(t) \rangle$ is defined as:

$$G(\tau) = \frac{\langle \delta F(t) \cdot \delta F(t+\tau) \rangle}{\langle F(t) \rangle^2} \tag{5.1}$$

Where, τ is the correlation time between a pair of photons detected at time t and $t + \tau$. $G(\tau)$ contains information about the molecular nature of the processes influencing the fluorescence intensity $F(t)$: diffusion, chemical kinetics, photo-physics and brightness. Under optimal excitation intensity, if only free diffusion is considered and the observation volume is an ellipsoidal 3D-Gaussian function, the theoretical autocorrelation $G_D(\tau)$ for a diffusive process is:

$$G_D(\tau) = \frac{1}{N} \cdot \left[1 + \frac{\tau}{\tau_{DT}}\right]^{-1} \cdot \left[1 + \frac{\tau}{\kappa^2 \cdot \tau_{DT}}\right]^{-\frac{1}{2}} \tag{5.2}$$

Here, N is the average number of molecules within the observation volume, τ_{DT} is the dwell time and κ is the axial ratio of the observation volume. The axial ratio $\kappa = \frac{z_0}{\omega_0}$, also called ellipticity, is the ratio between the major z_0 and minor ω_0 axis of the ellipsoidal observation volume and defines its shape. If the intersystem crossing transition $S_1 \rightarrow T_1$ (Fig.4.2) is also considered, a multiplicative term must be added to the diffusive autocorrelation function $G_D(\tau)$:

$$G_T(\tau) = \frac{1 - A_T + A_T \cdot e^{-\frac{\tau}{\tau_T}}}{1 - A_T} \cdot G_D(\tau) \tag{5.3}$$

Where, A_T is the population of the dark state T_1 and τ_T is the characteristic time constant of the intersystem crossing kinetic transition. The two model functions, $G_D(\tau)$ and $G_T(\tau)$, are normally used to analyze typical FCS measurements in the $\mu s - ms$ time domain.

5.3 Determination of N

In real experiments, the calculated autocorrelation data $G(\tau_i)$ are typically fitted with one of the theoretical behaviors described by Eq.5.2 and Eq.5.3 (see Fig.5.1).

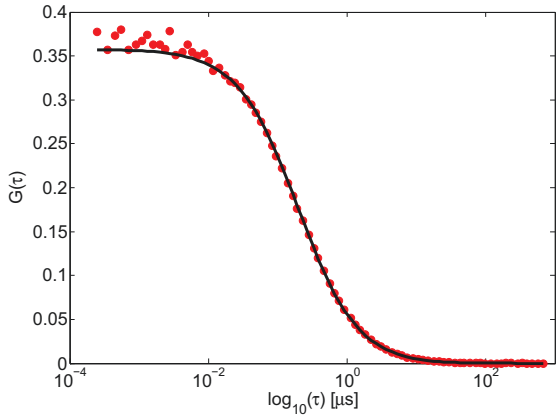


Figure 5.1: Typical experimental autocorrelation function $G(\tau_i)$ fitted with a purely diffusive model $G_D(\tau)$ (Eq.5.2).

Here, the autocorrelation function at time zero G(0) is used to calculate an important experimental parameter N_{FCS}, which is the apparent average number of molecules inside the observation volume. Depending on the model, Eq.5.4 or Eq.5.5 is used:

$$N_{FCS} = \frac{1}{G_D(0)} \tag{5.4}$$

$$N_{FCS} = \frac{1}{G_T(0) \cdot (1 - A_T)} \tag{5.5}$$

The measured N_{FCS} values do not always correspond to the average number of molecules N because of the background fluorescence bg_f (see Fig.5.2a). The theoretical relation between N and N_{FCS} in the presence of background noise can be derived if the detected number of photons within a correlation time τ is low: $(f_f \cdot \tau) << 1$ [42]. Since at correlation time $\tau = 0$ this condition is always satisfied, the desired relation can be written:

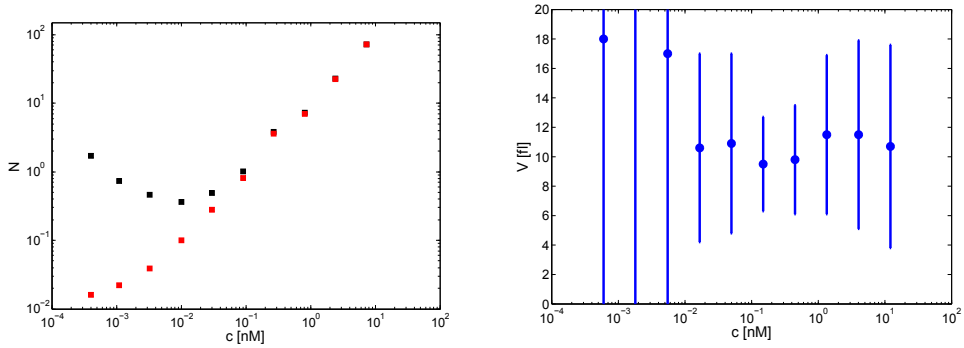
$$N = N_{FCS} \cdot \frac{(N \cdot f_f)^2}{(N \cdot f_f + bg_f)^2} \tag{5.6}$$

This equation shows that at fixed excitation intensity, at which the molecular brightness f_f is constant (see Eq.4.2), the deviation of the measured value N_{FCS} from the true value N depends only on the sample concentration $(N = c \cdot V)$. Eq.5.6 together with Eq.4.6 can be

used to calculate N from the experimentally determined values of N_{FCS}, s_f, and bg_f:

$$N = N_{FCS} \cdot \frac{(s_f - bg_f)^2}{s_f^2} \tag{5.7}$$

From the knowledge of N, the molecular brightness f_f can eventually be estimated, as described in Sec.4.1.1. In addition, the best concentration range to perform FCS measurements can be determined if a concentration series is measured (Fig.5.2). Typically, N values between 2 and 5 are optimal in FCS; in fact, at lower N, the background noise dominates (Fig.5.2a), and, at higher N, the amplitude of the autocorrelation function drops to zero (Eq.5.4 and Eq.5.5). In both cases, the errors of the estimated fitting parameters increase, as shown in Fig.5.2b for the observation volume V.



(a) (b)

Figure 5.2: Atto655-NHS concentration series: (a) Measured N_{FCS} values in black compared with the background corrected N values in red; (b) Observation volume V estimated from the fit of the FCS autocorrelation curves.

5.4 Determination of D, R_g and R_h

The main contribution to fluorescence intensity fluctuations $\delta F(t)$ in solution is the translational diffusion of the molecules crossing over the observation volume. For freely diffusing molecules, the diffusion coefficient D is defined from the values of ω_0 and τ_{DT}:

$$D = \frac{\omega_0^2}{4 \cdot \tau_{DT}} \tag{5.8}$$

As can be easily imagined, the diffusion coefficient encloses information about the microscopic

nature of the diffusing molecule, like dimension and shape, as well as the properties of the solvent.

A simple physical model [48], derived by Einstein and Stokes for spherical rigid particles of radius R_a diffusing in a continuos solvent, is generally used to describe diffusive processes in solution. Typically, diffusing macromolecules like proteins do not have a spherical shape. In addition, these molecules are quite flexible and show hydrodynamic behavior. Therefore, an effective hydrodynamic sphere must be defined to extend the model validity to proteins. The radius of this sphere is called hydrodynamic radius R_h. The Stokes-Einstein equation states that:

$$D = \frac{k_B \cdot T}{6\pi \cdot \eta(T) \cdot R_h} \tag{5.9}$$

Here, k_B is the Boltzman's constant $1.38065 \times 10^{-23} \frac{m^2 \cdot Kg}{s \cdot K}$, T is the temperature in [K] and $\eta(T)$ is the temperature dependent viscosity of the solvent $[Pa \cdot s]$. Eq.5.9 permits to compute the hydrodynamic radius R_h [Å] from the experimentally determined diffusion coefficient D $\left[\frac{m^2}{s}\right]$. Since in Eq.5.9 an ideal spherical protein is considered, the hydrodynamic radius does not give any information about the protein shape and size. Furthermore, it encloses electrostatic forces and solvation effects. Therefore, diverse empirical correlations have been proposed along the years to link the diffusion coefficient with the protein size and shape [49]. Here, the linear correlation between D and the radius of gyration R_g proposed by Tyn et al. was used [48]:

$$D = 5.78 \times 10^{-8} \cdot \frac{T}{\eta(T) \cdot R_g} \tag{5.10}$$

Where, D is in $\left[\frac{cm^2}{s}\right]$, T is in [K], η in [cP] and R_g in [Å]. Here, the constant 6.85×10^{-15}, which has the dimension of $\left[\frac{cm^2 \cdot cP \cdot Å}{s \cdot K}\right]$, was obtained by fitting 198 data points from 86 proteins. Eq.5.10 shows that a reshape of the protein structure will reflect itself in a change of the diffusion coefficient throughout the radius of gyration R_g. Thereby, large conformational movements of the protein can be monitored by means of the measured D.

5.5 Precise FCS Measurements

As shown in Sec.5.3 and Sec.5.4, the parameters of interest from a FCS measurement are: the average number of molecules N and the diffusion coefficient D. To estimate D with confidence, the dimension of the observation volume V must be known, and the sample concentration must be properly selected to have an optimal signal-to-noise ratio (see Sec.5.3). By definition the

ellipsoidal observation volume V is:

$$V = \pi^{\frac{3}{2}} \cdot \omega_0 \cdot \kappa \tag{5.11}$$

The axial ratio κ specifies the shape of the detection volume, whereas, the minor axis ω_0, its absolute dimension. With this definition in hand, Eq.5.8 can be redefined, in order to perform FCS calibration measurements with reference samples with known diffusion coefficients D:

$$D = \left(\frac{V}{2\pi^{\frac{3}{2}} \cdot \kappa}\right)^2 \cdot \frac{1}{\tau_{DT}} \tag{5.12}$$

The diffusion coefficient D is now a function of the volume parameters (V and κ) and the dwell time τ_{DT}. The values of κ and τ_{DT} can be retrieved from the fit of the measured autocorrelation function $G(\tau_i)$ of the reference sample and fixed. Thereby, the volume dimension V can be tuned in Eq.5.12 until the diffusion coefficient D matches the expected value. The calibrated volume dimensions, V and κ, can finally be used to determine the diffusion coefficients of unknown samples by FCS measurements. Here, if the experiments are not carried out at the reference temperature of 25 °C, the expected diffusion coefficients must be corrected for the temperature dependence [35].

In this thesis, Atto655-NHS and Atto488-NHS dyes were employed to calibrate the blue and red observation volume, respectively, because their diffusion coefficients in water are well known from dual focus FCS measurements [50][51]. To execute the calibration, the two reference dyes were dissolved in Milli-Q water. The concentration was set to N between 2 and 5 to have an optimal signal-to-noise ratio (Sec.5.3), and the solution was kept at room temperature to avoid artefacts due to temperature mismatches. The laser excitation intensity I_e was set to values lower than the saturation intensity I_s (Sec.4.2), but within the linear regime. Before every experiment, the observation volume was calibrated with the reference dyes. The values for V and κ, averaged over 4 months of measurements, are displayed in Tab.5.1.

Table 5.1: Diffusion coefficients of the reference dyes measured with two focus FCS at 25°C in water [50][51]. Measured detection volumes V and ellipticities κ averaged over 4 months of measurements for 470 nm and 640 nm laser excitations.

reference dye	λ_{ex}[nm]	$D^{H_2O}_{25\,^\circ C}[\frac{\mu m^2}{s}]$	V[fl]	κ
Atto488-NHS	470	369	3.8 ± 0.6	10.9 ± 1.6
Atto655-NHS	640	425	7.9 ± 1.6	7.8 ± 0.7

As indicated by the standard deviation, the measuring system is quite stable and does not require frequent realignement. Nevertheless, the dimension of the observation volumes was periodically cross-checked with a concentration series measurement and by imaging of sub-

resolution fluorescent beads, as described in [46].

5.6 Summary

In this chapter, the basics of fluorescence correlation spectroscopy, for standard measurements in the μs – ms time interval, were given. The two models used to fit the data in this time regime were introduced with Eq.5.2 and Eq.5.3. The effect of background noise on the estimated average number of molecule N was discussed in Sec.5.3, and the experimental correction factor derived in Eq.5.7. The corrected N values are used in Sec.4.1.1 to calculate the molecular brightness f_f. The diffusion coefficient D can be retrieved from FCS experiments, and the radius of gyration R_g calculated with Eq.5.10, revealing the microscopic properties of the diffusing molecule. Protein conformational changes can be monitored through it (see Sec.10.5). The optimal conditions to perform a precise FCS measurements are reported in Sec.5.5. A sample concentration corresponding to N between 2 and 5 is required. Furthermore, the calibration procedure to determine the observation volume V with a reference sample is described in Sec.5.5.

Chapter 6

Fluorescence Lifetime

6.1 Introduction

Abstracted from [52][53]

The fluorescence lifetime τ measures the average time spent by a fluorescent molecule in the excited state prior to its returning to the ground state. In the course of these few nanoseconds, a lot of interesting processes can affect the lifetime. Owing to the lifetime, fluorophores are sensitive reporters about the microscopic environment in which they are embedded. By virtue of the lifetime, a fluorescent dye gives an account of the solvent properties, and finally, because of the lifetime, two fluorophores reveal their relative distance up to $\sim 100\,\text{Å}$. Lifetime can record kinetic processes with a characteristic time τ_c lower than the lifetime itself. Static distributions of states are distinguished through the lifetime. A correct lifetime estimate is needed to measure anisotropy. As a result of this all, the lifetime is perhaps the most important parameters to be measured in fluorescence spectroscopy. A simple definition of lifetime is given in this chapter to connect all the possible microscopic contribution to the measured lifetime. The methods to measure lifetime and to get a proper lifetime estimate at ensemble and single-molecule level are discussed.

6.2 Microscopic Origin of the Fluorescence Lifetime

By definition, the fluorescence lifetime τ is the average time spent by a fluorophore in the excited state before decaying, radiatively or not, to the ground state. From a microscopic point of view, the fluorescence lifetime arises when a laser pulse excites one electron to a higher energy level. Its subsequent decay to the ground-state reveals the nature of the fluorescing molecule coupled to the environment. A proper description of the possible de-excitation mechanisms sheds light on the thread connecting the molecular processes with the experimentally

determined for a donor
dye in the p

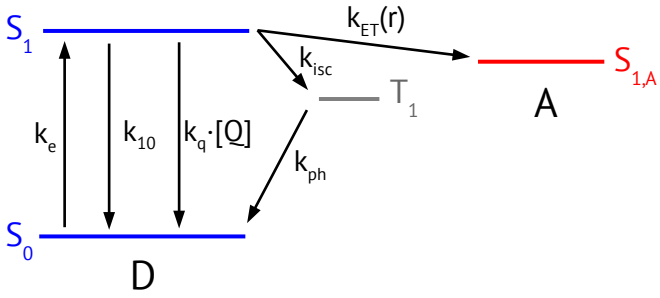


Figure 6.1: Three state model for a donor dye (D) in the presence of the acceptor (A). The radiative (k_{10}) and not radiative (k_q, k_{ISC} and $k_{ET}(r)$) de-excitation rates are shown together with the excitation rate k_e.

At first the acceptor is omitted, thence, once in the excited state S_1, the electron can decay radiatively to the ground state $S_1 \xrightarrow{k_{10}} S_0$. Alternatively, the excited state can be deactivated without photon emission, either by the intersystem crossing towards the triplet state $S_1 \xrightarrow{k_{isc}} T_1$, or, via a bimolecular reaction $S_1 \xrightarrow{k_q \cdot [Q]} S_0$, by quenchers Q freely diffusing in solution. All these relaxation processes may affect the fluorescence lifetime and are described by microscopic rates constants k $\left[\frac{1}{s}\right]$. In general, the transition rate k_{isc} towards the triplet-state is smaller compared to the other rates and can be discarded. Thus, the fluorescence lifetime τ_{D0} of an isolated donor fluorophore is:

$$\tau_{D0} \simeq \frac{1}{k_{10} + k_q \cdot [Q]} \tag{6.1}$$

Here, k_{10} is the rate of radiative decay towards the ground state S_0, and k_q $\left[\frac{1}{M \cdot s}\right]$ is the bimolecular quenching rate. From the rate k_{10} the natural or intrinsic lifetime of the fluorophore $\tau_n = \frac{1}{k_{10}}$ is defined. The rate of radiative decay k_{10} is sensitive to the environment surrounding the fluorophore including the solvent (Stokes shift). On the other hand, the quenching rate k_q depends on the exposure of the fluorescent dye to the quenchers Q, like tryptophan fluorescence for instance.

If the acceptor dye is also considered, the description must include the energy transfer rate $k_{ET}(r) = \frac{1}{\tau_{D0}} \cdot \left(\frac{R_0}{r}\right)^6$ from the donor excited state towards the acceptor excited state $S_1 \xrightarrow{k_{ET}(r)} S_{1,A}$ (Fig.6.1). With this new contribution, the fluorescence lifetime of the donor coupled to the acceptor $\tau_{DA}(r)$ is:

$$\tau_{DA}(r) \simeq \tau_{D0} \cdot \left[1 + \frac{R_0}{r}\right]^{-6} \tag{6.2}$$

Since the rate of energy transfer $k_{ET}(r)$ is a function of the Förster radius R_0 and the inter-dye

distance r, the measured donor lifetime $\tau_{DA}(r)$ reports about the relative distance r between the donor and the acceptor dye. The shorter the distance the lower the lifetime and vice versa. For that to happen, the behavior of the donor only sample must be characterized by measuring τ_{D0} (see Sec.9.4). The Förster radius R_0 is a property of the FRET pair (Sec.8.2); it depends on the strength of the dipolar coupling and defines the characteristic length scale of the interaction. Typically, inter-dye distances between 20 Å and 100 Å can be probed. Eq.6.1 and Eq.6.2 are the foundations to elucidate the microscopic physical meaning of the measured lifetimes, and are required to build up a reasonable model for data interpretation.

6.3 Ensemble Fluorescence Lifetime Measurements

Ensemble fluorescence lifetime measurements were performed as described in [35]. The fluorescence intensity decay I(t) of the fluorophore was recorded on the confocal microscope upon excitation with a pulsed laser at 20 MHz. An optimal laser excitation intensity $I_e = I_{opt}$ was used to avoid optical saturation (Sec.4.2). The experiments were performed in Time Correlated Single Photon Counting (TCSPC) mode (Sec.7.2). Solutions with N_{FCS} between 2 and 5 were measured. The aquisition time was adjusted in order to have $\sim 10^5$ counts at the maximum peak position. The typical lifetime intensity decays I(t) were fitted in MATLAB with a proper model function F(t), after iterative reconvolution with the measured Instrument Response Function - IRF(t). The weighted least-squares residuals (WLS), calculated with the Neyman's approach [54], were minimized with FMINUIT [36]. The goodness of fit was assessed from the χ^2-distribution [54]. A $1\,\%$ significance level was tolerated. The two model functions $F_{ME}(t)$ and $F'_{DA}(t)$ used in this thesis are introduced in the next subsections.

6.3.1 Multiexponential Model

The most straightforward model $F_{ME}(t)$ used to fit fluorescence data I(t) is a discrete weighted sum of single-exponential decaying functions. This model, because of its symplicity and powerfulness, is often abused, and can lead to overfitting and misinterpretation of the data. The model function $F_{ME}(t)$ is:

$$F_{ME}(t) = \sum_{i=1}^{n} a_i \cdot e^{-\frac{t}{\tau_i}} \tag{6.3}$$

The fitting parameters are: the amplitudes of the exponential components at time zero a_i, and the decay times τ_i. A multiexponential decay $F_{ME}(t)$ can originate from a mixture of different fluorophores or from the same dye surrounded by diverse environments. In both cases, the resolved decay times τ_i are generated by different radiative k_{10} or quenching k_q rates.
From the fitting parameters (a_i, τ_i), the fractional contribution f_i of each decay to the total

integrated intensity $\int I(t)\,dt$, and the intensity averaged lifetime $\langle\tau\rangle_f$ can be calculated:

$$f_i = \frac{a_i \cdot \tau_i}{\sum a_i \cdot \tau_i} \tag{6.4}$$

$$\langle\tau\rangle_f = \sum f_i \cdot \tau_i = \frac{\sum a_i \cdot \tau_i^2}{\sum a_i \cdot \tau_i} \tag{6.5}$$

The fluorescence intensity fraction f_i defines the percentage of photons belonging to one decay with respect to the total number of emitted photons. In addition, the amplitude fraction x_i and the amplitude average lifetime $\langle\tau\rangle_x$ can be defined:

$$x_i = \frac{a_i}{\sum a_i} \tag{6.6}$$

$$\langle\tau\rangle_x = \sum x_i \cdot \tau_i = \frac{\sum a_i \cdot \tau_i}{\sum a_i} \tag{6.7}$$

The meaning of the fraction x_i depends on the physical origin of the different time decays τ_i. If the decay stems from a mixture of dyes, each fraction x_i is related to the relative concentration of the individual species. Conversely, if every component generates from the same flurophore in different states, x_i represents the population of each conformer. Because of the different weights, the amplitude averaged lifetime is always lower than the intensity averaged lifetime ($\langle\tau\rangle_x < \langle\tau\rangle_f$), except for a single-exponential decay where they coincide. The model function $F_{ME}(t)$ adequately describes the behavior of free dyes or dyes bound to macromolecules. The amplitude average lifetime $\langle\tau\rangle_x$ is used in Sec.9.3.3 to calculate the quantum yields of the dyes bound to PGK and DNA. The intensity average lifetime $\langle\tau\rangle_f$ of the donor only species is employed in Sec.8.6 as a parameter to build the 2D-plots. Tab.6.1 displays the fitting parameters obtained from the decays of the free dyes used in this thesis.

Table 6.1: Fluorescence lifetime fitting parameters of free dyes. The lifetimes are reported in [ns].

sample	n_{exp}	$\langle\tau\rangle_f$	$\langle\tau\rangle_x$	τ_1	τ_2	x_1	f_1	χ^2
Al647-NHS	2	0.96	0.87	1.06	0.45	0.70	0.84	1.098
Nile Blue	2	1.40	1.35	1.43	0.45	0.92	0.97	1.118
Atto655-NHS	2	1.79	1.72	1.86	0.86	0.86	0.93	1.162
Fluorescine	1	3.98	/	/	/	1	1	1.137
Al488-NHS	1	4.04	/	/	/	1	1	1.053
Atto488-NHS	1	4.20	/	/	/	1	1	1.064

The blue dyes are well described by a single-exponential decay. Conversely, the red dyes show a double-exponential decay. This can be explained by two conformers with different rates of radiative decay k_{10}. The number of exponential components of the Alexa dyes increases when

they bind to DNA and PGK, because of local quenching effects, as shown in Tab.9.3.

6.3.2 Model for a Gaussian Distribution of Distances

Under many circumstances, the intensity decay originates from a continuum of states. The best example is the donor decay under FRET coupling. Here, the donor lifetime $\tau_{DA}(r)$ depends on the inter-dye distance (Eq.6.2), and, if the dye pair visits different states on timescales longer than the lifetime, a distribution of distances p(r) must be introduced. As a consequence, the recorded lifetime decay I(t) is an integral over the single exponential decays of each state weighted by a probability distribution of distances ($\int p(r)\,dt = 1$). The model function for the donor decay $F_{DA}(t)$ under FRET coupling is defined assuming an empirical Gaussian distribution of distances $p_G(r)$ and a donor only multiexponential decay $F_{D0}(t)$ (Eq.6.3):

$$F_{DA}(t) = A \cdot \sum_{i=1}^{n} x_i \cdot \int p_G(r) \cdot e^{-\left[\frac{t}{\tau_{D0,i}} \cdot \left(1+\frac{R_0}{r}\right)^6\right]} dr \tag{6.8}$$

$$p_G(r) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot e^{-\frac{(r-\langle r\rangle)^2}{2\sigma^2}} \tag{6.9}$$

The amplitude fractions x_i and the lifetimes $\tau_{D0,i}$ are obtained from the fit of the donor only decay, and fixed together with the Förster radius R_0. The free parameters are the amplitude A, and the mean $\langle r\rangle$ and the width σ of the Gaussian distribution. Physically speaking, the distribution p(r) is a probability density function and encloses the information about the conformational space sampled by the two dyes. Consequently, the mean $\langle r\rangle$ and the width σ indicate the average relative position of the dyes and the extent of the fluctuations around the mean $\delta r = r - \langle r\rangle$, respectively. As discussed in Sec.8.6.2, two mechanisms can induce such a distribution, the dye linker dynamics or the dynamics of the macromolecule.
In real experiments, a fraction of donor only molecule $F_{D0}(t)$ (Eq.6.3) and a background fluorescence component $F_{bg}(t)$ are also present. The background is described by a single-exponential decay $F_{bg}(t) = a_{bg} \cdot e^{-\frac{t}{\tau_{bg}}}$ and the total model function $F'_{DA}(t)$ is:

$$F'_{DA}(t) = a_D \cdot [(1 - x_{D0}) \cdot F_{DA}(t) + x_{D0} \cdot F_{D0}(t)] + a_{bg} \cdot e^{-\frac{t}{\tau_{bg}}} \tag{6.10}$$

Here, x_{D0} is the fraction of donor only species, and a_D and $a_{bg} = 1 - a_D$ are the amplitudes of the dye and the background fluorescence components, respectively. An important application of the model is the correction of the static line for inter-dye distance fluctuations faster than milliseconds (Sec.8.6.2). The model function $F'_{DA}(t)$ is also used in Sec10.3 to fit the donor intensity decays of dlPGK in order to study the protein interdomain dynamics.

6.4 Single-Molecule Lifetime Estimator

In single-molecule experiments, a molecule enters the detection volume, randomly diffuses inside for a few milliseconds, and, finally, leaves the volume. During this little amount of time, only a few tens of photons are recorded. For this reason, a well defined intensity decay I(t) cannot be measured, and the full complexity of the lifetime decay cannot be revealed from the data. Nevertheless, the fluorescence lifetime estimated at the single-molecule level is a valuable tool to reveal photo-physical and dynamic effects by means of smFRET experiments (Chap.8).

The Maximum Likelihood Estimator (MLE) is generally used to calculate the lifetime from single-molecule data τ_{sm} [55][56][57]. In this thesis, a simpler method is used [53]. The single molecule lifetime τ_{sm} is calculated as the arithmetic mean of the delay times δt_i of the photons recorded in the course of a single-molecule transit inside the observation volume (Appx.B.1). In fact, by definition, the lifetime is the average time $\langle t \rangle$ spent by the fluorophore in the excited state following the excitation. If the delay time δt_i is defined as the time interval between the laser excitation at $t_0 = 0$ and the i-photon emitted at time t_i ($\delta t_i = t_i - t_0$), the single-molecule lifetime estimator τ_{sm} is the arithmetic mean:

$$\tau_{sm} = \frac{1}{N_{ph}} \sum_{i=1}^{N_{ph}} \delta t_i \tag{6.11}$$

Where, N_{ph} is the number of recorded photons for each single-molecule event. In Eq.6.11, the information about the microscopic behavior of the probe is averaged out and partially lost. In fact, single- and multiexponential decays, as well as more complex decays, can generate the same single-molecule lifetime value τ_{sm}. Furthermore, it must be kept in mind that the lifetime estimated from single-molecule experiments τ_{sm} is an intensity averaged lifetime $\langle \tau \rangle_f$ (see Appx.B.1):

$$\tau_{sm} = \langle \tau \rangle_f \tag{6.12}$$

Eq.6.12 is a cornerstone in the interpretation of the single-molecule FRET experiment with 2D-plots [58].

6.5 Summary

In Sec.6.2, the basics to understand the microscopic origin of the fluorescence lifetime were given. The powerfull three state model was used to elucidate the possible de-excitation mechanisms contributing to the measured lifetime (Fig.6.1). The two fundamental model functions $F_{ME}(t)$ and $F'_{DA}(t)$, which were used to fit ensemble lifetime decays, were introduced in Sec.6.3.

The microscopic physical origin of the two models was also explained. The multiexponential model function $F_{ME}(t)$ (Sec.6.3.1) well describes free dyes or dyes bound to macromolecules, whereas, a Gaussian distribution of distances was introduced in Sec.6.3.2 to fit the donor decay under FRET coupling. The amplitude averaged lifetime $\langle\tau\rangle_x$ is going to be used in Sec.9.3.3 to calculate the quantum-yield of dyes bound to DNA and PGK. The intensity averaged lifetime $\langle\tau\rangle_f$ of the donor only sample is needed in Sec.10.4 to build the 2D-plots. The width σ of the Gaussian distance distribution will be employed in Sec.8.6.2 to correct the static line for inter-dye distance fluctuations that are faster then milliseconds. Finally, the method of choice to calculate the lifetime τ_{sm} from single-molecule measurements was given in Sec.6.4. The connection of τ_{sm} with the intensity averaged lifetime $\langle\tau\rangle_f$, that is estimated from ensemble experiments, was also shown in Eq.6.12. More detailed information about the derivation of the Eq.6.11 and Eq.6.12 are found in Appx.B, together with the methodology used to correct the bias of the estimated single-molecule lifetime values τ_{sm}.

Chapter 7

The Stream of Information in Single-Molecule Measurements

7.1 Introduction

Single-molecule fluorescence measurements were performed in this thesis with a confocal microscope on molecules freely diffusing in solution. Since FRET was the main application, this method is analyzed here. Easily speaking, a confocal microscope is a highly sensitive and highly precise stop-watch which measures photons. Thus, to extract valuable physical information from single-molecule fluorescence measurements, the photons must be recorded, stored, classified, and analyzed in a specific way. A stream of information emerges, when an isolated diffusing fluorophore excited by a laser pulse generates a typical fluorescence burst. The information starts flowing, when the recorded photons are filtered and classified depending on their energy, polarization, and origin. The stream gets to the most crucial step, when the fluorescence bursts are distinguished with respect to the background, and the fluorescence intensity and lifetime are recovered. Finally, the stream of information gets to the apex, when the physical parameters are calculated and analyzed, and the information reaches the user. In this chapter, the first three steps of the information flow are discussed, placing the focus on single-molecule FRET measurements. The last step is analyzed in the next chapter (Chap.8), and eventually examples are examined in the last two chapters (Chap.9 and Chap.10).

7.2 Photon Acquisition and Classification: PIE, TCSPC and T3

In this work, all the single-molecule FRET (smFRET) meaurements were performed with four Single Photons Avalanche Detectors (SPAD). The fluorescence emitted photons collected by

the objective were first separated by color with a dicroic mirror and then by polarization with a polarizer beam splitter. Finally, emission filters were used to clean the signal. Therefore, only this experimental set-up is considered here.

The stream of information starts, when a laser pulse excites the FRET sample freely diffusing inside the observation volume. Pulsed Interleaved Excitation (PIE) [59] was applied with a delay of 25 ns between the direct excitation of the donor and the acceptor, in order to collect all the emitted photons from one laser pulse before the next pulse with a different color arrives (Fig.7.1). The photons were recorded using Time-Correlated Single Photon Counting (TCSPC) [47]. Three tags: the absolute arrival time, the delay time and the detection channel were finally stored in the so called T3-mode data format [58]. The absolute arrival time t_M, also called macro time, is the interval between the k-laser pulse at time $t_{L,k} = k \cdot \Delta t_L$, which generated the photon with pulsing frequency $f_L = \Delta t_L^{-1}$, and the start of the experiment at time t_0. The resolution is on the nanoseconds timescale. The delay time t_m, also called micro time, is the delay between the excitation at time $t_{L,k}$ and the fluorescence emission at time t, and has picoseconds resolution. The detection channel N_{Ch} univocally connects one photon to one detector. Each of this parameters carries a different physical information.

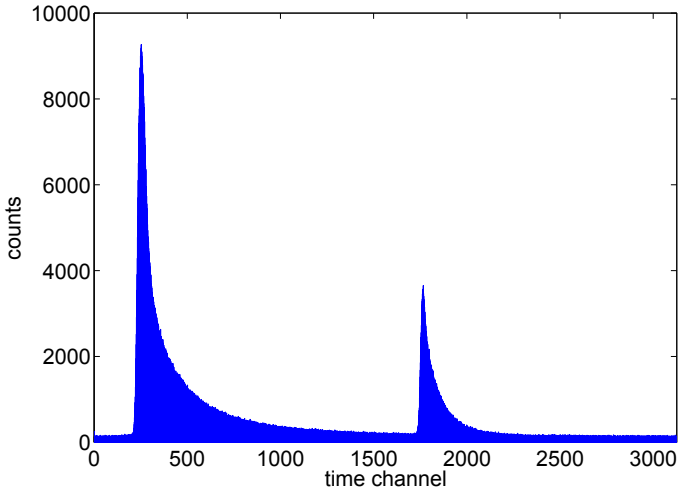


Figure 7.1: Typical TCSPC histogram under PIE excitation at 20 MHz. The red laser (640 nm) is delayed by 25 ns with respect to the blue laser (470 nm). The two spikes corresponds to the donor and acceptor direct excitation, respectively. A time channel corresponds here to $\Delta t_{Ch} = \frac{50\,ns}{3126} = 16\,ps$.

In the second step of the stream of information, the origin of the detected photons is totally recovered by a combination of the information obtained from PIE, TCSPC and the detection channel. First, the photons are divided into two groups depending on the excitation source.

Practically, two time windows are applied to the TCSPC histogram to separate photons generated by direct excitation of either the donor or the acceptor. For example, in Fig.7.1, the two time windows, from channel 200 to channel 1700 and from channel 1700 to channel 3126, correspond to photons emitted via direct excitation of either the donor or the acceptor, respectively. Second, the photons are classified respect to their polarization ($\perp$ or $\parallel$) and energy (blue or red) by using the detection channel information. Because of that, the configuration of the optical filters must be known. In the used set-up, Ch.1 and Ch.2 correspond to red parallel and red perpendicular photons, while Ch.3 and Ch.4 relate to the blue parallel and blue perpendicular quanta of light. With this last information, the origin of a single photon is totally recovered, and the next step of the stream of information can be tackled.

7.3 Burst Selection Criterion

The central problem in single-molecule fluorescence spectroscopy is the discrimination of the little amount of fluorescence photons F_f generated by a molecule crossing through the observation volume from the background BG_f. The approach described in [43] was used because it utilizes all the information encoded in the photon stream. Furthermore, the threshold criterion used for the single-molecule events discrimination is more stringent with respect to the methods based on binned time traces. In the approach used in this thesis, the information encoded in the sequence of recorded absolute times $\{t_{M,i}\}$ is employed to build up the the inter-photon distance traces: $IPD(t_i) = t_{M,i} - t_{M,i-1}$ (Fig.7.2). Since the inter-photon distance is inversely proportional to the total count rate s_f, the typical bursts of fluorescence generated by single-molecule events appear as drops in the $IPD(t_i)$ trace. A threshold criterion is then applied to the smoothed $IPD(t_i)$ trace to distinguish these events from the background, and the selected groups of photons are used to calculate the physical parameters of individual fluorescence bursts.

In this work, the threshold criterion was exclusively applied to the $IPD(t_i)$ trace of the directly excited acceptor, that is usually called PIE channel. Afterwards, the edges of individual bursts were employed to pick out the blue and red photons generated by direct excitation of the donor. As a consequence, the bursts generated by donor only molecules are automatically omitted, and the bursts are cut when the acceptor bleaches. A moving average filter [43][60] with half-width $m = 25$ was used to smooth the data, in order to preclude the distortion of the burst size distribution [61]. The threshold of the $IPD(t_i)$ trace of the PIE channel was set to $T_{IPD} = 160\,\mu s$ for optimally excited acceptor bound to dlDNA and dlPGK (Tab.4.2). This value was estimated from the knowledge of the measured fluorescence brightnesses f_f of the acceptor under direct excitation, which are displayed in Fig.A.1 and Fig.A.2, and the knowledge of the average background count rates $bg_f = (1 - 2.5)\,MHz$.

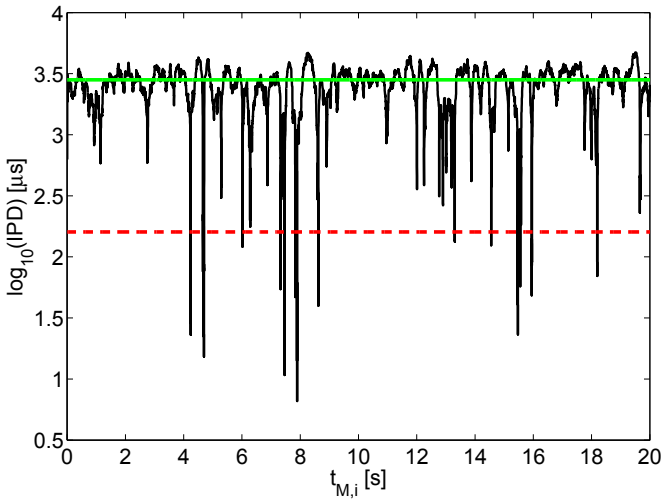


Figure 7.2: Inter-Photon Distance trace plotted on a logarithmic scale as a function of the macro time $t_{M,i}$ of each detected i-photon. Also the threshold $T_{IPD} = 160\,\mu s$ for the burst selection (red dashed) and the average background (green continuous) are displayed.

7.4 Multiparameter Fluorescence Detection - MFD

In the next step of the information flow, a set of physical parameters is calculated for each single-molecule event; from here arises the term Multiparameter Fluorescence Detection (MFD) [58][62]. At first, the fluorescence intensity and the lifetime are obtained burstwise for each group of photons depending on the color (B or R), the polarization ($\perp$ or $\parallel$), and origin (D, A or PIE). The fluorescence intensities F_f are calculated by substraction of the number of background photons BG_f from the total number of photons S_f detected in one burst. To do so, Eq.8.6 and Eq.8.7, which consider also the cross-talk α of the donor fluorescence into the red detection channel, were employed. Finally, the lifetimes are calculated with Eq.B.3 and Eq.B.8. Only bursts with a total intensity $N = F_D + F_A$ higher than a threshold N_T are considered for FRET analysis, in order to discard noisy bursts with very low number of photons. Threshold values of 25 and 30 were used for dlPGK and dlDNA, respectively. From the recovered intensities, further physical quantities like the steady state anisotropy r, the energy transfer efficiency E, and the ratio F' can be derived (Sec.8.3).

7.5 MATLAB Scripts

Two MATLAB scripts were used to perform the operations reported in the previous sections. The first script, provided by PicoQuant GmbH, was used to convert the T3 output file of the

MicroTime 200 machine to an -ASCII file. The script was self-modified by adding a loop in order to process more than one input file in a single run. The output of the script is a matrix with the channel number, the macro time, and the micro time of each detected photon.
The second script was self-implemented to classify the photons (Sec.7.2), select individual bursts (Sec.7.3), and calculate the intensity, the lifetime, and the duration of each burst (Sec.7.4). This second script loads the output file of the first script, and gives back an -ASCII matrix with the burst coordinates that are the set of physical parameters calculated for each burst. Also, the macro times and the micro times of the individual group of photons are saved in -ASCII format.
Further analysis of the physical parameters was performed with a third script, as described in Sec.8.7.

7.6 Frequency Histograms and Optimal Single-Molecule Experimental Conditions

A peculiar feature of single-molecule techniques is the possibility to access in its full complexity the distribution of states underlying the observed phenomenon. Single-molecule experiments are also unique to reveal the interconversion between states visited by the system during the observation time. Nevertheless, the last two statements are strictly true only if complete time averaging over the observation time [63][64] and multimolecule crossing events [65] are absent. The first requirement sets a limit to the lower characteristic time τ_c of the dynamics accessible by single-molecule experiments (see Sec.8.6). The second condition must be fulfilled to avoid artefacts and misinterpretation of the data.
Generally, single-molecule experiments are analyzed with frequency histograms of the physical parameters calculated for a set of fluorescence bursts. In order to construct well defined histograms, the statistics acquired in one burst and over the set of bursts must be good. In this work, an optimal excitation intensity (Sec.4.2) and a $100\,\mu$m pin-hole were used to get large number of photons in individual burst (see Eq.4.3). Sets of 10 to 12 times one hour measurements were performed to acquire a large amount of events $(15000 - 25000)$. Furthermore, the number of events was kept constant during the measurement by preventing sticking of the sample to the cover slide with Sigmacote®(Sec.2.5), and with a $0.003\,\%$ volume fraction of Tween®20 added to the measuring solution. Finally, a concentration of the red fluorophore corresponding to $N_{FCS} = 0.03$ was used. This value is a compromise between the optimal value of $N_{FCS} = 0.01$ [65] and the need to record a reasonable number of events during one day of measurements.

7.7 Summary

In this chapter, the firsts steps of the information flow, starting with the excitation of the molecule under the focus, and ending with the calculation of the interesting physical parameters, were analyzed. The PIE excitation scheme was introduced in Sec.7.2 together with the TCSPC acquisition method and the T3 data format. The powerfull combination of PIE, TCSPC and T3 allows one to classify the detected photons depending on their energy, polarization and origin, and to perform Multiparameter Fluorescence Detection (MFD) (Sec.7.4). The burst selection criterion based on the application of a threshold T_{IPD} to the Inter-Photon Distance trace $IPD(t_i)$ was described in Sec.7.3. Self-written MATLAB scripts were implemented to accomplish these tasks. Finally, the conditions to perfom single-molecule FRET experiments with good statistics and without artefacts are reported in Sec.7.6. All the recovered information is going to be used in Chap.8 to analyze the smFRET data.

Chapter 8

Analysis of smFRET Measurements with 1D and 2D Frequency Histograms

8.1 Introduction

In the last step of the information flow, the fluorescence intensities F_f and lifetimes τ_{sm} obtained with the single-molecule FRET experiments are used to extract information about the states distribution of the system, and about the system dynamics. The fluorescence intensities F_f are first transformed to distance dependent quantities, the FRET indicators (E and F'). Afterwards, the efficiency distributions p(E) are displayed with 1D frequency histograms, and then analyzed to explore the underlying distribution of the inter-dye distances p[r(t)]. Finally, the donor lifetime dimension τ_{DA} is added to the 1D efficiency histogram to investigate the dyes photo-physical properties, the dynamics of the system, and the states connectivity. The resulting two dimensional frequency histograms are called 2D plots. A theoretical relation, called static line, is used to emphasize how different contributions manifest themself in the 2D plots. Two empirical correlations are also introduced to describe fast and slow processes with respect to the burst duration timescale (ms). To open the chapter, a glance is given to the microscopic origin of the Förster Resonance Energy Transfer.

8.2 Förster Resonance Energy Transfer - FRET

Förster Resonance Energy Transfer (FRET) arises from the weak dipolar coupling between the excited states of two fluorescent molecules called donor (D) and acceptor (A) [66]. This process is characterized by the energy transfer rate $k_{ET}(r)$ and by the energy transfer efficiency

E, which defines the probability that the excited state of the donor transfers its energy to the acceptor rather than decaying to the ground state [53]. Therefore, by definition, the FRET efficiency of a FRET pair, which is described by the energy scheme displayed in Fig.6.1, is:

$$E(r) \equiv \frac{k_{ET}(r)}{k_{D0} + k_{ET}(r)} \tag{8.1}$$

Here, $k_{ET}(r)$ is the rate of energy transfer from the donor to the acceptor, and $k_{D0} = \frac{1}{\tau_{D0}}$ is the decay rate of the donor without the acceptor (Eq.6.1). The previous equation (Eq.8.1) shows that the transfer efficiency E(r) is controlled by the inter-dye distance (r), which modulates the aperture of the FRET channel $k_{ET}(r)$. An explicit equation for the efficiency E is derived, keeping in mind that $k_{ET}(r) = \frac{1}{\tau_{D0}} \cdot \left(\frac{R_0}{r}\right)^6$ (see Sec.6.2):

$$E(r) = \left[1 + \left(\frac{r}{R_0}\right)^6\right]^{-1} \tag{8.2}$$

Eq.8.2 shows how the transfer efficiency relies on the physical parameters of the system. The power to the sixth dependence on the inter-dye distance (r) derives from the dipole-dipole interaction, which mediates the energy transfer process, and explains the high sensitivity of the FRET measurements to small distance changes around the Förster radius R_0 (Fig.8.1).

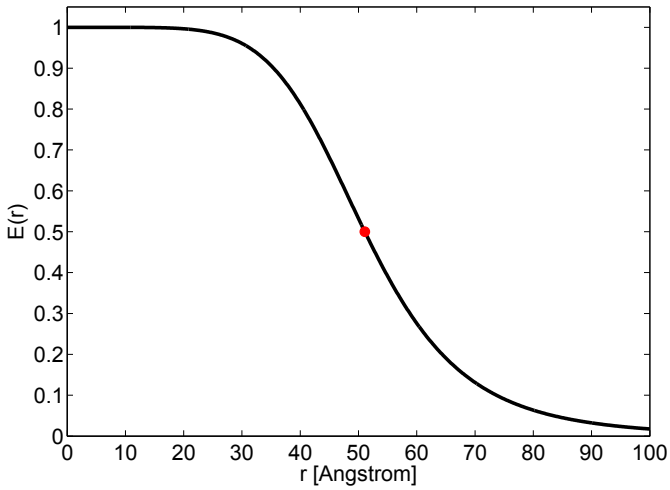


Figure 8.1: Inter-dye distance dependence of the energy transfer efficiency E(r) for the Al647/Al488 FRET pair bound to DNA. $R_0 = 51.1$ Å, $n = 1.334$, $\phi_D = 0.87$ and $\epsilon_A(\lambda_{max}) = 239000\,M^{-1}cm^{-1}$. The value of the Förster radius R_0 is displayed by a red circle.

On the other hand, the Förster radius R_0 encloses the dependence of the energy transfer efficiency on the dyes properties [52][67]. In fact, R_0 [Å] specifies the strength of the dipolar coupling between the donor and the acceptor, and depends on the dyes spectral and photophysical properties, on the relative orientation of the molecular dipoles, and on the solvent characteristics:

$$R_0 = 0.211 \cdot \left[J(\lambda) \cdot k^2 \cdot \phi_D \cdot n^{-4} \right]^{\frac{1}{6}} \tag{8.3}$$

Here, $J(\lambda) = \int f'_D(\lambda) \cdot \epsilon_A(\lambda) \cdot \lambda^4 \, d\lambda$ is the overlap integral between the corrected fluorescence emission spectrum of the donor $f'_D(\lambda)$ (Appx.D.3) normalized to its area and the exctinction spectrum of the acceptor $\epsilon_A(\lambda)$ $\left[\frac{nm^4}{M \cdot cm}\right]$; k^2 is the dyes orientation factor and is generally set to $\frac{2}{3}$ assuming free rotation of the dyes; ϕ_D is the quantum yield of the donor, and n is the refractive index of the solution. R_0 determines the length scale of the dipolar interaction and has typical values between 30 Å and 60 Å, which means that distances between 20 Å and 100 Å can be monitored with nanoscopic resolution (Fig.8.1).

8.3 Single-Molecule FRET Indicators: E and F'

The definitions of energy transfer efficiency given in Eq.8.1 and Eq.8.2 explain the microscopic origin of the energy transfer process, but the FRET indicators must be derived to analyze the single-molecule experiments. A first FRET indicator, the energy transfer efficiency E, is obtained as shown in Appx.C.1:

$$E = \left[1 + \frac{f_D}{f_A} \right]^{-1} = \frac{F_A}{F_A + \gamma \cdot F_D} \tag{8.4}$$

Where, f_D and f_A are the fluorescence detection rates, F_D and F_A the measured fluorescence intensities, and $\gamma = \gamma' \cdot g'$ is a correction factor product of two ratios, the quantum yield ratio γ' and the detection efficiency ratio g'. A second FRET indicator F', which is the ratio between the fluorescence emission rate of the donor k_D and the acceptor k_A, is also derived as a function of the fluorescence detection rates f_f and the fluorescence intensities F_f [58]:

$$F' \equiv \frac{k_D}{k_A} = \frac{f_D}{f_A} \cdot g' = \frac{F_D}{F_A} \cdot g' \tag{8.5}$$

In this work, the ratio $\gamma' = \frac{\phi_A}{\phi_D}$ was calculated from the measured quantum yields (Sec.9.3.3), and the detection efficiency ratio $g' = \frac{g_A}{g_D}$ was obtained with a calibration measurement performed on double-labeled DNA (Sec.9.7). Then, the above mentioned equations (Eq.8.4 and

Eq.8.5) were used to calculate the FRET efficiency E and the ratio F' of individual bursts of fluorescence. In order to do so, the detected fluorescence intensities F_f were calculated from the measured total S_f and background BG_f intensities (Eq.4.5), taking into account the cross-talk α (Sec.9.5) of the donor fluorescence into the acceptor detection channel. The following expressions were used:

$$F_D = S_D - BG_D \tag{8.6}$$

$$F_A = S_A - BG_A - \alpha \cdot F_D \tag{8.7}$$

The transfer efficiencies E obtained from the fluorescence bursts were used to build up the 1D frequency histograms which were analyzed with an empirical model function (Sec.8.5). For qualitative survey of the data with 2D plots, the F' ratio was used because of the higher resolution with respect to the efficiency E (Sec.8.6) [58].

8.4 Broadening of the p(E) and p(F') Distributions

The primary information attained from single-molecule experiments are the distributions of the FRET indicators p(E) and p(F'). These distributions are obtained from the data by constructing 1D frequency histograms of the measured $\{E_i\}$ and $\{F'_i\}$ sets of values. To understand the physical origin of these distributions, a further relation, which explains the connection between the underlying physical quantities and the FRET indicators E and F', must be introduced (Appx.C.2):

$$\frac{f_D}{f_A} = \frac{r^6 \cdot n^4}{\phi_A \cdot g' \cdot k^2 \cdot J(\lambda)} \tag{8.8}$$

This last expression, as shown by Eq.8.4 and Eq.8.5, describes how the FRET indicators depend on a set of physical parameters: $\{\xi_i\} = \{r, n, \phi_A, g', k, J(\lambda)\}$. In principle, each parameter ξ_i may have its own distribution of values $p(\xi_i)$ with mean $\langle\xi_i\rangle \equiv \int \xi_i \cdot p(\xi_i)\, d\xi$. Consequently, in the most general case, the variance $\sigma^2_{E/F'}(\{\xi_i\})$ of the FRET indicators distributions p(E) and p(F') is a complex function of the set of parameters $\{\xi_i\}$. For instance, assuming to have independent variables $\{\xi_i\}$, the total variance $\sigma^2_{E/F'}(\{\xi_i\})$ is:

$$\sigma^2_{E/F'}(\{\xi_i\}) = \sigma^2_{SN} + \sum_i \sigma^2_{E/F'}(\xi_i) \tag{8.9}$$

Where, the shot noise variance σ^2_{SN} add up with the terms dependent on the fluctuations of the parameters $\{\xi_i\}$. For example, $\sigma^2_E(\xi_i) \equiv \langle E(\xi_i)^2\rangle - \langle E(\xi_i)\rangle^2$ is the fraction of the transfer efficiency variance $\sigma^2_E(\{\xi_i\})$ dependent on ξ_i. Eq.8.9 shows that when all the distributions $\{p(\xi_i)\}$ are completely averaged out during the observation time T, which means that $\sigma^2_{E/F'}(\xi_i) = 0\ \forall_i$,

the distributions p(E) and p(F') of the FRET indicators are broadened only by the shot noise ($\sigma^2_{E/F'} = \sigma^2_{SN}$). Furthermore, under the same hypothesis, the mean values $\langle E \rangle$ and $\langle F' \rangle$ are determined exclusively by the mean $\langle \xi_i \rangle$ of the individual parameters. Conversely, if at least one of the distribution of the set $\{p(\xi_i)\}$ is not averaged out during the observation time, the variance $\sigma^2_{E/F'}(\xi_i)$ becomes a function of the parameter ξ_i. Therefore, the measured p(E) and p(F') distributions reflect the properties of the underlying distributions $\{p(\xi_i)\}$.
In general, in single-molecule experiments, all the distributions $\{p(\xi_i)\}$ are averaged out during the observation time T of a few milliseconds except the inter-dye distance distribution p[r(t)] and the acceptor quantum yield distribution $p(\phi_A)$. The second term $p(\phi_A)$ is explained by local quenching of the acceptor fluorescence in different labeling positions or by slow interconversion between different conformers [68]. On the other hand, the inter-dye distance distribution p[r(t)] depends on the equilibrium distribution of states of the system and on the system dynamics. Consequentely, it encloses the interesting physical information. Told that, the total variance $\sigma^2_E(\{\xi_i\})$ of the transfer efficiency distribution p(E) is approximated as:

$$\sigma^2_E(\phi_A, r) \simeq \sigma^2_{SN} + \sigma^2_{\phi_A} + \sigma^2_r(T) \tag{8.10}$$

Explicit relations for each term of Eq.8.10 are found in Appx.C.3. Here, it is worth remembering that, the conformational variance $\sigma^2_r(T)$ reflects itself on the measured p(E) distribution as a function of the characteristic time τ_c of the inter-dye distance fluctuations with respect to the observation time T (Appx.C.3.3) [63]. When the dynamics are faster than the burst duration ($\tau_c << T$), the distance distribution is averaged out during the observation time ($\sigma^2_r(T) = 0$), and the single-molecule experiments do not confer more information with respect to ensemble measurements. Conversely, if the dynamics timescales are comparable to the burst duration ($\tau_c \sim T$), information about the distribution of states, the states connectivity and the dynamics timescales can be obtained. An example is given in Sec.10.4.

8.5 1D FRET Efficiency Histograms

The FRET efficiency distributions p(E) recovered from single-molecule experiments are normally analyzed with different methodologies [69][70][71]. In the present work, the efficiency distributions p(E) were displayed with 1D frequency histograms built up with the measured efficiencies values $\{E_i\}$. The histograms were then fitted with an empirical multi-gaussian model function G(E):

$$G(E) = \sum_{i=1}^{m} \frac{p_i}{\sqrt{2\pi}\sigma_i^2} \cdot e^{-\frac{(E-\langle E_i \rangle)^2}{2\sigma_i^2}} \tag{8.11}$$

Where, after selecting the number of gaussians (i), the fitting parameters are: the variance σ_i^2, the average FRET efficiency $\langle E \rangle_i$, and the amplitude p_i of each population. The mean values $\langle E \rangle_i$ were used to calculate the FRET averaged inter-dye distance $\langle r_{DA} \rangle_{E_i}$ of each population with Eq.8.2. On the other hand, the FRET variances $\sigma_{E_i}(\phi_A, r)^2 = \sigma_i^2$ (Eq.8.10) were used to assess the presence of dynamic interconversion occuring during the burst duration. This was done by comparing the measured variances σ_i^2 with the sum of σ_{SN}^2 and $\sigma_{\phi_A}^2$, calculated as described in Appx.C.3. The extra-terms were attributed to conformational variance $\sigma_r^2(T)$.

8.6 2D Plots (F' vs. τ_{DA}): Static Line

In the 2D-plots, the information from the single-molecule fluorescence lifetimes and from the FRET estimators are combined. In this work, the lifetime of the donor in the presence of the acceptor τ_{DA} and the ratio F' were combined to build up the 2D frequency histograms. The expected correlation between τ_{DA} and F' for a quasi-static distribution of distances $p[r(t)] \sim p(r)$ is defined by the static line (Fig.8.3) [72][58]:

$$F' = \frac{1}{\gamma'} \cdot \frac{\tau_{DA}}{\langle \tau_{D0} \rangle_f - \tau_{DA}} \tag{8.12}$$

Here, γ' is the ratio of the acceptor and donor quantum yields defined in Sec.8.5, and $\langle \tau_{D0} \rangle_f$ is the intensity averaged lifetime of the donor without the acceptor. These two parameters were measured as described in Sec.9.3.3 and Sec.9.4. Deviations of the experimental populations from the static line indicate either quenching of the dyes or dynamic populations $p[r(t)]$ with distance fluctuations slower than the lifetime and faster than the observation time ($\langle \tau_{D0} \rangle_f < \tau_c \leq T$). On the other side, populations lying on the static line represent quasi-static conformations $p[r(t)] \sim p(r)$. Furthermore, the shape of the experimental 2D populations reveals the timescales of the dynamics with respect to the burst duration (T). This allows one to distinguish between fast ($\tau_c << T$) and slow ($\tau_c \sim T$) dynamics on the milliseconds timescale, as well as to assess the states connectivity with a qualitative survey of the data (Fig.8.3).

8.6.1 Dye Quenching

Quenching of fluorescence appears in the 2D-plots of F' versus τ_{DA} with a shift of the population above the static line. Two quenching processes can take place: static or dynamic. Dynamic quenching reduces the quantum yield and the lifetime of the dye, while, static quenching decreases only the quantum yield [52]. Therefore, static quenching of the donor is not visible in the 2D-plots because F' (Eq.8.8) and τ_{DA} are unaffected, while dynamic quenching of the donor reduces the donor lifetime τ_{DA} shifting the population to the left with respect to the

static line. Talking about the acceptor, static or dynamic quenching increase the F' ratio (Eq.8.8) moving the population above to the static line. In this thesis, the parameters γ' and $\langle\tau_{D0}\rangle_f$ used to calculate the static and the dynamic lines (Eq.8.12) were corrected for quenching of the donor and the acceptor, as shown in Appx.D.4.5.

8.6.2 Fast Dynamics: "Corrected" Static Line

Fast dynamics refers to distance fluctuations slower than the donor lifetime $\langle\tau_{D0}\rangle_f$ and much faster than the burst duration T ($\langle\tau_{D0}\rangle_f < \tau_c << T$). Fast dynamics appears in the 2D-plots with well defined populations below the static line (Fig.8.3a). The effects of fast distance fluctuations can be accounted for in the 2D-plots with an empirical polynomial transformation applied to the static line [72]:

$$\langle\tau_{DA}\rangle_x = \sum_{i=0}^{n} c_i \cdot \langle\tau_{DA}\rangle_f^i \tag{8.13}$$

The polynomial coefficients $\{c_i\}$ are obtained by fittting the $\{\langle\tau\rangle_{x,j}\}$ and $\{\langle\tau\rangle_{f,j}\}$ values calculated as described in Appx.C.4. An example of the fitting procedure is shown in Fig.8.2. The transformed amplitude averaged lifetimes values $\langle\tau_{DA}\rangle_x$ calculated with Eq.8.13 are then plugged in Eq.8.12 to calculate the "corrected" static line (Fig.8.3a).

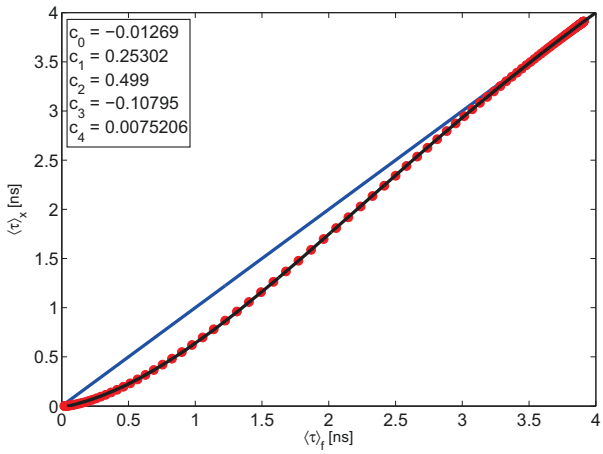
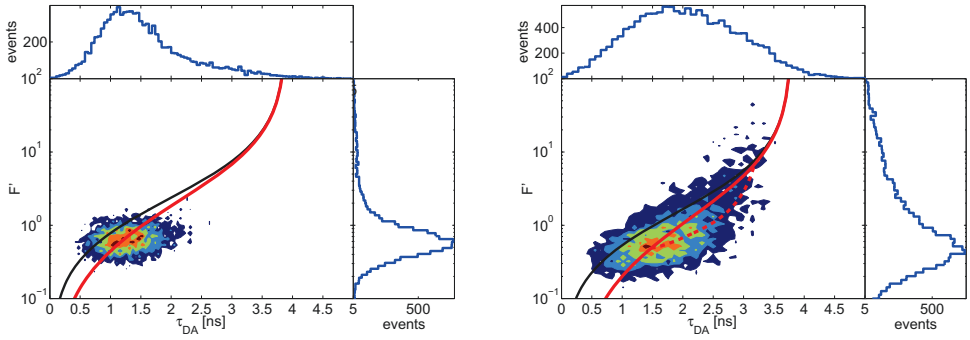


Figure 8.2: Example of fitted $\{\langle\tau\rangle_{x,j}\}$ versus $\{\langle\tau\rangle_{f,j}\}$ values calculated with the following parameters: $\sigma_{DA} = 6$ Å and $\langle\tau_{D0}\rangle_f = 3.91$ ns. The blue line represents the expected behavior for a static distribution of distances $p[r(t)] \sim p(r)$. The red circles were calculated as described in Appx.C.4 and the black line stands for the 5th degree polynomial fit. The coefficients $\{c_i\}$ of the fit are shown in the figure caption. The "corrected" static line is shown in Fig.8.3a with a red continuous line.

In this thesis, the static lines were always corrected for fast dynamics interconversion, to highlight the presence of additional slower contributions. A common source of fast distance fluctuations for dyes bound to proteins is the linker dynamics, which is the diffusion of the fluorophore that takes place on the ~ 100 ns [73] timescale within the volume delimited by the protein surface and the length of the linker. However, even protein motions can tune fast fluctuations of the inter-dye distances. In both situations, the distance distribution p[r(t)] is well described by a gaussian function $p_G(r)$, which can be attained by fitting the ensemble TCSPC (eTCSPC) histogram of the donor in the double-labeled sample, as described in Sec.6.3.2. Thus, if the two dynamics are superimposed and independent, the single terms in a first approximation are additive [74], and the total variance $\sigma^2_{DA}(r)$ is:

$$\sigma^2_{DA}(r) = \sigma^2_{LD}(r) + \sigma^2_{PD}(r) \tag{8.14}$$

Where, $\sigma^2_{LD}(r)$ and $\sigma^2_{PD}(r)$ are the linker and protein contributions to the dynamics, respectively. Eq.8.14 shows that, without additional information, the individual terms cannot be decoupled from the measured total variance $\sigma^2_{DA}(r)$.



(a) Fast Dynamics **(b)** Slow Dynamics + Fast Dynamics

Figure 8.3: 2D-plots for: (a) Fast linker dynamics and (b) Fast dynamics with slow inter-conversion between 2-states. The static line (black continuos), the "corrected" static line (red continuos) and the 2-states dynamic line (red dashed) are displayed. These correlations were calculated with Eq.8.12. For the "corrected" static line and the 2-state dynamic line the polynomial transformation shown in Fig.8.2 and Fig.8.4 were used, respectively. The two examples were taken from Sec.9.7.2 and Sec.10.4.2.

8.6.3 Slow Dynamics: 2-State Dynamic Line

For dynamics comparable to the observation time ($\tau_c \sim T$), single-molecule FRET experiments provide more information with respect to the ensemble measurements. In fact, the underlying

distribution of states p[r(t)] does not average out during the observation time. Under these circumstances, the experimental populations are smeared out and lie below the static line, therefore revealing the connectivity of the interconverting states and their relative populations (Fig.8.3b). Since the fast dynamics of the linker are in general superimposed on the slow motions of the protein, only this situation is analyzed here. An empirical correlation shows the connectivity of the interconverting states and is called dynamic line (Fig.8.3b). This relation is obtained with the same procedure utilized for the "corrected" static line (Sec.8.6.2). The only difference resides in the way the set of amplitude $\{\langle\tau\rangle_{x,j}\}$ and intensity $\{\langle\tau\rangle_{f,j}\}$ averaged lifetimes are calculated (Appx.C.4.2). An example of the fitting procedure is displayed in Fig.8.4 for fast dynamics superimposed on slow interconversion between two states.

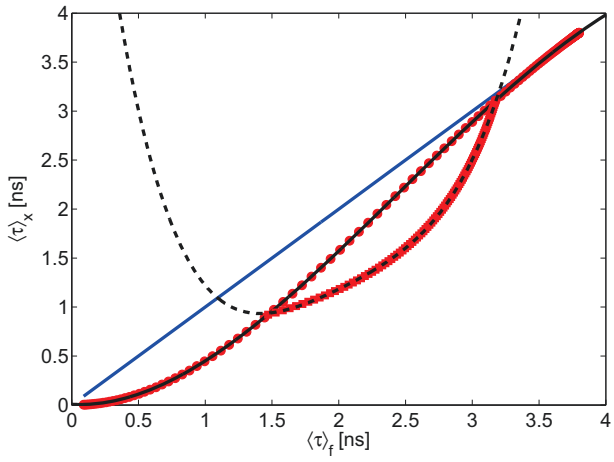


Figure 8.4: Example of fitted $\{\langle\tau\rangle_{x,j}\}$ versus $\{\langle\tau\rangle_{f,j}\}$ values calculated for fast dynamics (red circles) and slow dynamics (red squares) using: $\sigma^2_{DA} = 8.8$ Å, $\langle\tau_{D0}\rangle_f = 3.80$ ns, $R_0 = 51.0$ Å, $\langle r_{DA}\rangle_{E1} = 40.5$ Å and $\langle r_{DA}\rangle_{E2} = 68.1$ Å. The coefficients $\{c_i\}$ of the two polynomial fits are not shown. The resulting "corrected" static line and dynamic line are shown in Fig.8.3b.

8.7 MATLAB Scripts

A script was implemented to perform the analysis of the smFRET data. The inputs of the script are the ASCII file of the burst coordinates (Sec.7.5) and the parameters needed for the analysis (see Chap.9). Here, after the calculation of the FRET indicators (E and F') (Sec.8.3), the user can choose to build up and to fit the 1D efficiency histogram with the multi-gaussian model function (Sec.8.5). Thereafter, the 2D-plot is calculated and the user has the option to draw the "corrected" static line and/or the dynamic line (Sec.8.6). Finally,

the input parameters used for the analysis and the parameters obtained from it are saved in an ASCII file.

8.8 Summary

In this chapter, the microscopic origin of FRET was analyzed, and two equations (Eq.8.4 and Eq.8.5) were given to calculate the FRET indicators (E and F') from the fluorescence intensities F_f measured by the single-molecule experiments. The connection between the measured p(E) and p(F') distributions and the underlying distributions $\{p(\xi_i)\}$ of the physical parameters of the system was highlighted in Sec.8.4. Here, the focus was put on the inter-dye distance ditribution p[r(t)]. Then, the efficiency distribution p(E) was analyzed in Sec.8.5 with a multigaussian model function to investigate the state distributions. The FRET averaged distances $\langle r_{DA} \rangle_E$ and the variances σ_E^2 of each population are the main parameters obtained from this analysis. Later on, the donor lifetime information τ_{DA} was combined with F' in the 2D-plots, to sort out dye quenching effects, and to explore the dynamic properties of the underlying states distribution (Sec.8.6). From the 2D-plots, the timescale of the dynamics τ_c with respect to the burst duration T and the states connectivity can be retrieved. In order to do so, a theoretical relation called static line and two empirical correlations were used to describe the two different situations (Eq.8.13).

Chapter 9

Measurements of ϕ, τ_{D0}, α, R_0 and g'

9.1 Introduction

The 1D FRET efficiency histograms and the 2D plots described in the previous chapter are powerfull tools for the analysis of single-molecule FRET experiments. However, they rely on the knowledge of a set of physical parameters, which must be measured with high precision in order to get trustworthy information about the state distribution and the dynamics of the system. These physical quantities are: the donor and acceptor quantum yields (ϕ_D and ϕ_A), the intensity and the amplitude averaged lifetimes of the donor without the acceptor ($\langle\tau_{D0}\rangle_f$ and $\langle\tau_{D0}\rangle_x$), the cross-talk α, the Förster radius R_0, and the detection efficiency ratio g'. All these parameters are measured in this chapter, and then used for the analysis of the interdomain dynamics of phosphoglycerate kinase from yeast (Chap.10).

9.2 General Strategy for the Parameters Estimation

In this work, the FRET experiments were performed on PGK and DNA which were double-labeled with the acceptor (Al647-mal) and the donor (Al488-mal) (Chap.3). Therefore, the set of parameters $\{\xi\} = \{\phi, \tau, \alpha, R_0\}$ mentioned in the introduction were estimated for these samples with the following general strategies.

9.2.1 PGK

The double-mutant of phosphoglycerate kinase was labeled on the two different domains at position Q135C and S290C (Sec.3.1). Here, the donor and the acceptor are distributed between the two labeling sites with different probabilities (Sec.3.1.2). Therefore, in order to estimate the set of physical parameters $\{\xi\}$ of the dyes bound to double-labeled PGK, the relative population ρ_i of each site must be considered, and used to weight the values measured for

individual sites. In this work, Eq.3.4 was applied, and the physical quantities $\langle\xi\rangle_{\text{dlPGK}}$ were estimated for double-labeled PGK from the values measured on the single-labeled samples (ξ_{Q135C} and ξ_{S290C}):

$$\langle\xi\rangle_{\text{dlPGK}} = (0.36 \cdot \xi_{\text{Q135C}}) + (0.64 \cdot \xi_{\text{S290C}}) \tag{9.1}$$

9.2.2 DNA

Two double-labeled DNA samples, where the labeling positions are separated by 10bp and 17bp respectively, were used to measure the detection efficiency ratio g' (Sec.9.7). Here, the donor and the acceptor bind a specific site on a specific single-strand (see Sec.2.1.5). Consequentely, the properties of the dyes on the double-labeled and on the single-labeled species coincide ($\xi^{\text{dl}} = \xi^{\text{sl}}$). Furthermore, it turns out that the dye properties are independent on the labeling positions ($\xi_{\text{10bp}} \simeq \xi_{\text{17bp}}$). Therefore, the physical quantities ($\xi^{\text{sl}}_{\text{10bp}}$ and $\xi^{\text{sl}}_{\text{17bp}}$) measured with the two single-labeled samples were averaged and used to analyze the double-labeled DNA samples:

$$\bar{\xi}_{\text{dlDNA}} = \frac{\xi^{\text{sl}}_{\text{10bp}} + \xi^{\text{sl}}_{\text{17bp}}}{2} \tag{9.2}$$

9.3 Fluorescence Quantum Yield: ϕ

The quantum yields of the dyes attached to the double-labeled samples must be measured in order to analyze the smFRET experiments. In fact, the ratio of the quantum yields $\gamma' = \frac{\phi_{\text{A}}}{\phi_{\text{D}}}$ is necessary to calculate the set of FRET efficiencies $\{E_i\}$ (Eq.8.4), and the static and the dynamic lines (Eq.8.12). In addition, the donor quantum yield enters in the calculation of the Förster radius R_0 (Sec.9.6)

9.3.1 The Working Equations

The most confortable way to determine the fluorescence quantum yield ϕ is through a comparative method, which compares a selected physical property $\xi(\phi)$, that depends on the quantum yield of the unknown (U), with the same property measured for the reference (R). The following general working equation is used to calculate the unknown quantum yield ϕ_{U} of a dye with the comparative method:

$$\phi_{\text{U}} = B \cdot \frac{\xi_{\text{U}}(\phi)}{\xi_{\text{R}}(\phi)} \cdot \phi_{\text{R}} \tag{9.3}$$

Where ϕ_{R} is the quantum yield of the reference dye, $\xi(\phi)$ is a measurable physical property of the dye which depends on the quantum yield ϕ, and B is a measurable constant. Therefore,

in order to determine the unknown quantum yield ϕ_U with Eq.9.3, the value of the reference quantum yield ϕ_R must be known with confidence, and the functional dependence $\xi(\phi)$ must be explicitly written. In this thesis, three different expressions, $\tau(\phi)$, $f_f(\phi)$ and $F(\phi)$ were employed to derive the working equations (Eq.9.7, Eq.9.9 and Eq.9.11), which were applied to measure the unknown quantum yields ϕ_U.

a) Lifetime: $\xi(\phi) = \tau(\phi)$

The first expression $\xi(\phi)$ relates the quantum yield ϕ to the fluorescence lifetime $\tau(\phi)$. This expression is derived from the definition of quantum yield:

$$\phi \equiv \frac{k_{10}}{k_{10} + k_{nr}} \tag{9.4}$$

Eq.9.4 shows that the fluorescence quantum yield ϕ is the fraction of absorbed photons ($k_{10} + k_{nr}$) which are re-emitted via the radiative channel k_{10} (Fig.6.1). In Eq.9.4, the term $k_{nr} \simeq k_q \cdot [Q] + k_{ET}(r)$ encloses all the non radiative contributions. The desired functional dependence $\tau(\phi)$ is derived from Eq.9.4 by using the definition of lifetime $\tau \equiv \frac{1}{k_{10}+k_{nr}}$ (Eq.6.1), and by defining the natural lifetime of the fluorophore $\tau_n \equiv \frac{1}{k_{10}}$:

$$\tau(\phi) \equiv \tau_n \cdot \phi \tag{9.5}$$

Then, the working equation is derived from the lifetimes ratio $\frac{\tau_R(\phi)}{\tau_U(\phi)}$ considering two dyes with similar natural lifetimes ($\tau_{n,R} \simeq \tau_{n,U}$):

$$\phi_U^\tau \simeq \frac{\tau_U(\phi)}{\tau_R(\phi)} \cdot \phi_R \tag{9.6}$$

In real experiments, where dyes are bound to biomolecules, a discrete distribution of states with different decay rates (k_{10} and/or k_{nr}) is generally observed. Therefore, by definition, even the lifetime and the quantum yield (Eq.9.4) are distributed over a range of values $p(\tau)$ and $p(\phi)$, with means $\langle\tau\rangle_x$ and $\langle\phi\rangle$. As a consequence, Eq.9.5 becomes $\langle\tau(\phi)\rangle_x \equiv \tau_n \cdot \langle\phi\rangle$ (see Appx.D.2), and the working equation (Eq.9.6) generalizes to:

$$\langle\phi_U^\tau\rangle \simeq \frac{\langle\tau_U(\phi)\rangle_x}{\langle\tau_R(\phi)\rangle_x} \cdot \langle\phi_R\rangle \tag{9.7}$$

This equation was used to determine the unknown average quantum yield $\langle\phi_U^\tau\rangle$ from the measured amplitude averaged lifetimes ($\langle\tau_U\rangle_x$ and $\langle\tau_R\rangle_x$) obtained with a multiexponential fit of the intensity decays (Sec.6.3.1). In the prosecution, for simplicity, the averaging notation $\langle..\rangle$ will be discarded ($\phi^\tau = \langle\phi^\tau\rangle$) and it will be recalled just when needed for the comprehension.

Here, it is worth to point out again that only dyes with similar natural lifetimes ($\tau_{n,R} \simeq \tau_{n,U}$) can be compared, which means that a dye bound to a biomolecule must be compared with the same dye freely diffusing in solution. Finally, it is important to restate that the quantum yields ϕ_τ determined with the lifetime method are sensitive to dynamic quenching while they are not affected by static quenching. In fact, the lifetime is not affected by static quenching [52].

b) Molecular Brightness: $\xi(\phi) = f_f(\phi)$

The second expression $\xi(\phi)$ relates the quantum yield ϕ to the fluorescence detection rate or molecular brightness $f_f(\phi)$ measured on the confocal microscope at excitation intensities I_e below the saturation regime ($I_e < I_s$):

$$f_f(\phi) = [g_f \cdot \sigma_{01}(\lambda_{ex}) \cdot I_e] \cdot \phi \tag{9.8}$$

The microscopic origin of this expression is described in Sec.4.1.1. The working equation is derived by taking the brigthnesses ratio $\frac{f_R(\phi)}{f_U(\phi)}$, by considering similar excitation intensities ($I_U \simeq I_R$), and by using the proportionality between the absorption cross-section and the normalized optical density $\left(\sigma_{01}(\lambda) = \frac{OD'(\lambda)}{3.82 \cdot 10^{-21}}\right)$ [52]:

$$\phi_U^f \simeq \frac{g_{R,opt} \cdot OD'_R(\lambda_{ex})}{g_{U,opt} \cdot OD'_U(\lambda_{ex})} \cdot \frac{f_U(\phi)}{f_R(\phi)} \cdot \phi_R \tag{9.9}$$

Where $OD'(\lambda) = \frac{OD(\lambda)}{OD(\lambda_{max})}$ is the normalized optical density. Here, as a first approximation the detection efficiency of the photodetectors g_{el} is considered to be almost constant for the unknown and the reference ($g_{U,el} \simeq g_{R,el}$). Therefore, the ratio between the total detection efficiencies $\frac{g_R}{g_U} = \frac{g_{R,el} \cdot g_{R,opt}}{g_{U,el} \cdot g_{U,opt}} \simeq \frac{g_{R,opt}}{g_{U,opt}}$ depends only on the measurable optical detection efficiency g_{opt}, which depends on the optical elements of the microscope (Sec.4.1.1). Then, all the quantities in Eq.9.9 are measurable, and the working equation can be used to determine the unknown quantum yield ϕ_U^f from the measured molecular brightnesses $f_U(\phi)$ and $f_R(\phi)$. The quantum yields ϕ_f determined by the brigthness method are sensitive both to dynamic and to static quenching. In fact, the brightness depends on both quenching effects [52].

c) Integrated Fluorescence Intensity: $\xi(\phi) = F(\phi)$

The third and last expression $\xi(\phi)$ relates the quantum yield ϕ to the integrated fluorescence intensity $F(\phi)$ measured on the spectrofluorimeter, assumed low optical densities ($OD_{max} \leq 0.01$) [75]:

$$F(\phi) \propto \frac{OD(\lambda_{ex}) \cdot I(\lambda_{ex})}{n^2} \cdot \phi \tag{9.10}$$

Where $F(\phi)$ is the integral of the corrected emission spectrum $f'(\lambda)$ (Appx.D.3), $OD(\lambda_{ex})$ is the optical density at the excitation wavelength λ_{ex}, $I(\lambda_{ex})$ is the excitation intensity, and n is the refractive index of the solution. Eq.9.10 shows that only a fraction ϕ of the absorbed photons $N_{abs} = OD(\lambda_{ex}) \cdot I(\lambda_{ex})$ is re-emitted via the radiative channel. The refractive index n compensates for refraction at the cuvette interfaces [75]. The working equation is derived by taking the fluorescence intensities ratio $\frac{F_R(\phi)}{F_U(\phi)}$, and by rearranging the terms in the same form of Eq.9.3. Comparable ($I(\lambda_U) \simeq I(\lambda_R)$) and monochromatic ($BP_{ex} = (1-2)\,nm$) excitation intensities are also considered for the reference and the unknown in order to simplify the expression:

$$\phi_U^F \simeq \frac{n_U^2 \cdot OD_R(\lambda_{ex})}{n_R^2 \cdot OD_U(\lambda_{ex})} \cdot \frac{F_U(\phi)}{F_R(\phi)} \cdot \phi_R \tag{9.11}$$

This equation can be exploited to calculate the unknown quantum yield ϕ_U^F in two ways. In the first case, the concentration of the reference and unknown samples are adjusted in order to have similar optical densities at the excitation wavelength $OD_R(\lambda_{ex}) \simeq OD_U(\lambda_{ex})$. In fact, under this condition Eq.9.11 is simplified, and ϕ_U^F can be calculated from the measured $\frac{F_U(\phi)}{F_R(\phi)}$ and $\frac{n_U^2}{n_R^2}$ ratios. In the second case, Eq.9.11 is applied in a differential form to increase the precision of the measurement:

$$\phi_U^F \simeq \frac{n_U^2}{n_R^2} \cdot \frac{\partial F_U(\phi)}{\partial OD_U(\lambda_{ex})} \cdot \frac{\partial OD_R(\lambda_{ex})}{\partial F_R(\phi)} \cdot \phi_R \tag{9.12}$$

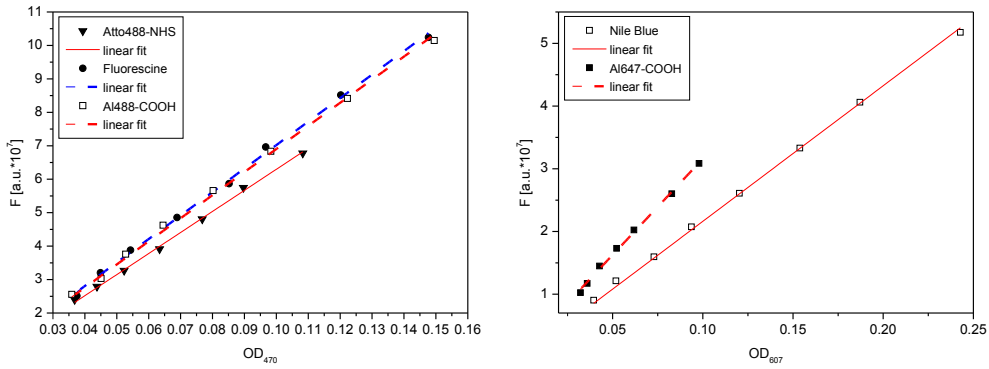
Here, since at low optical density a linear relation (Eq.9.10) holds between the integrated intensities and the optical densities ($F(\phi) \propto OD(\lambda_{ex})$), the derivatives can be replaced by the linear slopes $m_{U/R}(\phi) = \frac{\partial F_{U/R}(\phi)}{\partial OD_{U/R}(\lambda_{ex})}$ in order to obtain the final relation:

$$\phi_U^F \simeq \frac{n_U^2}{n_R^2} \cdot \frac{m_U(\phi)}{m_R(\phi)} \cdot \phi_R \tag{9.13}$$

This expression was used to cross-calibrate the quantum yields ϕ_R of the reference dyes with a concentration series measurement (Sec.9.3.2).

9.3.2 Cross-Calibration of the Reference Dyes

As shown by Eq.9.3, quantum yield measurements with the comparative methods require stable and reliable reference fluorophores with known quantum yields ϕ_R. Furthermore, for simplicity, the reference and the unknown should emit in the same spectral range, or at best being the same fluorophore bound and unbound to the biomolecule. Therefore, the reference dyes must be selected with care, and the reference quantum yields ϕ_R must be cross-calibrated prior to use to assure the reliability of the measurements.



(a) Blue Dyes **(b)** Red Dyes

Figure 9.1: Linear fits with zero intercept of the $\{OD_i(\lambda_{ex}), F_i\}$ data measured with a concentration series.

In this work, Eq.9.13 was used to perform the cross-calibration. Here, the data obtained with a concentration series were fitted with a linear relation in order to get the linear slopes m (Fig.9.1). Before going on with the discussion, it is worth to note that the corrected emission spectra $f'(\lambda)$ are the main source of uncertainity of the unknown quantum yields ϕ_U^F determined with this method. Relative errors of $\pm 10\,\%$ can be easily expected [52][76][75].
In this work, unbound Al488-COOH and Al647-COOH were choosen as the reference fluorophores because they are photo-physically stable in water solutions, and because the dyes bound to PGK and DNA were Al488 and Al647. Moreover, before performing the quantum yield measurements on the confocal microscope, the two reference dyes Al647-COOH and Al488-COOH were cross-calibrated with Nile Blue, and with Fluorescine and Atto488-NHS, respectively. The results of the cross-calibration are reported in Tab.9.1. The percentage relative deviations ($\delta\phi\,\% = 100 \cdot (\phi_U^F - \phi_R)$) between the measured ϕ_U^F and the expected ϕ_R values are comparable with the uncertainities of the corrected emission spectra $f'(\lambda)$ (see Fig.D.2b), which are the limiting factor of the measurement. Therefore, unbound Al488-COOH and Al647-COOH can be safely used as the reference samples for the quantum yield determination (Sec.9.3.3). Furthermore, the amplitude averaged lifetimes $\langle\tau_R\rangle_x$, which were measured the same day and on the same samples employed for the cross-calibration, were used to check the stability of the reference dyes immediately before the quantum yield measurements.

9.3.3 Quantum Yield Ratio: γ'

The donor and acceptor quantum yields (ϕ_D and ϕ_A) of the double-labeled samples (PGK and DNA) must be estimated in order to calculate the quantum yield ratio $\gamma' = \frac{\phi_A}{\phi_D}$. In this work,

Table 9.1: For each dye used for the cross-calibration are reported the expected quantum yields ϕ_R at 22 °C in the reference buffers. For Nile Blue, the reported quantum yield ϕ_R was corrected for its strong temperature dependence $\left(\frac{\Delta\phi}{\Delta T} = -0.0043\,°C^{-1}\right)$ [77]. The measured slopes m and the refractive indexes n were used to cross-calibrate the expected values ϕ_R of the dyes emitting in blue (Al488-COOH, Atto488-NHS, Fluorescine) and in red (Al647-COOH, Nile Blue). The resulting percentage relative deviations $\delta\phi\,\% = 100 \cdot (\phi_U^F - \phi_R)$ are reported. The amplitude averaged lifetimes $\langle\tau_R\rangle_x$ corresponding to the reference quantum yields ϕ_R were also measured.

sample	reference buffer	ϕ_R	$\langle\tau_R\rangle_x$ [ns]	n	m	$\delta\phi_1\,(\%)$	$\delta\phi_2\,(\%)$
Al488-COOH	PBS	0.92†	4.04	1.3355	69.06	1.4 %	3.8 %
Atto488-NHS	H_2O	0.80‡	4.20	1.333	62.97	1.4 %	2.5 %
Fluorescine	0.1 M NaOH	0.92§	3.98	1.334	70.21	4.4 %	2.5 %
Al647-COOH	PBS	0.33†	0.87	1.3355	30.66	7.6 %	/
Nile Blue	EtOH (0.5 % $\frac{v}{v}$ HCl)	0.261*	1.35	1.3635	21.62	6.9 %	/

† Invitrogen website, ‡ Atto-Tec product sheet, § [78] and * [77].

the quantum yield of Al488 and Al647 bound to double-labeled DNA and PGK were measured on the confocal microscope by means of two independent comparative methods, the "lifetime method" and the "brightness method", which are based on Eq.9.7 and Eq.9.9, respectively. In this way, the comparison of the measured ϕ_τ and ϕ_f quantum yields allows one to discern the presence or absence of static quenching. In fact, static quenching would reduce ϕ_f but not ϕ_τ. Furthermore, the precision of the measurements is increased. The whole procedure is described in detail in Appx.D.4, and the best estimates are reported in Tab.9.2, together with the calculated quantum yield ratios γ', which were then used to analyze the smFRET data (Sec.9.7 and Sec.10.4).

Table 9.2: Measured quantum yields ϕ_U and quantum yield ratios γ' of double-labeled DNA and PGK. See Appx.D.4 for the full derivation.

sample	ϕ_U	$\gamma' = \frac{\phi_A}{\phi_D}$
A-dlDNA	0.39	0.44
D-dlDNA	0.87	/
A-dlPGK	0.54§	0.67
D-dlPGK	0.81§	/

§ Corrected for the effects of PGK's substrates.

9.4 Donor Only Fluorescence Lifetime: $\langle\tau_{D0}\rangle_f$

The lifetimes of the donor only samples must be measured, in order to fit the donor intensity decay of the double-labeled samples with a gaussian distribution of distances $p_G(r)$ (Sec.6.3.2), and in order to calculate the static and the dynamic lines (Sec.8.6). The fluorescence lifetimes of the donor only samples, slPGK (Q135C and S290C) and slDNA ("10bp" and "17bp"), were measured at ensemble level, as described in Sec.6.3, and were fitted with a double exponential model function (Sec.6.3.1). In addition, a set of measurements was performed on slPGK with and without the substrates (Mg_2ATP, KADP, 3-PG). Afterwards, the expected donor only lifetimes were estimated with Eq.9.1 for double-labeled PGK, in order to consider the distribution of the dyes between the two labeling sites. The results are reported in Tab.9.3. From the measured intensity averaged lifetimes $\langle\tau_{D0}\rangle_f$, it is evident that Mg_2ATP and KADP quench the donor fluorescence, while 3-PG increases the donor lifetime. On the other hand, the two dlDNA samples have equivalent intensity averaged lifetimes (~ 3.91 ns).

Table 9.3: Decay times $\tau_{D0,i}$ and amplitude fractions x_i of donor only dlDNA and dlPGK. The intensity averaged lifetimes $\langle\tau_{D0}\rangle_f$ are also reported. For dlPGK the reported values are a weighted average of the values measured with the single-labeled samples (Eq.9.1). Different substrates were added to the PGK buffer solution (Mg_2ATP and KADP at 13 mM, 3-PG at 42 mM).

sample	$\tau_{D0,1}$ [ns]	$\tau_{D0,2}$ [ns]	x_1	x_2	$\langle\tau_{D0}\rangle_f$ [ns]
dlDNA 10bp	1.42	4.03	0.12	0.88	3.91
dlDNA 17bp	1.57	4.06	0.15	0.85	3.90
PGK + substrates	$\tau_{D0,1}$ [ns]	$\tau_{D0,2}$ [ns]	x_1	x_2	$\langle\tau_{D0}\rangle_f$ [ns]
none	1.65	3.98	0.17	0.83	3.80
Mg_2ATP	1.46	3.88	0.17	0.83	3.71
KADP	1.52	3.90	0.16	0.84	3.74
3-PG	1.93	4.05	0.15	0.85	3.91
3-PG*KADP	1.70	3.94	0.17	0.83	3.75

9.5 Cross-Talk: α

The cross-talk α must be measured in order to calculate the acceptor fluorescence intensities $\{F_{A,i}\}$ of a set of single-molecule events (Eq.8.7). The cross-talk is the fraction of photons emitted by the donor that are sensed by the acceptor detection channel. Therefore, by definition, and by considering that $f_{DD} >> f_{DA}$:

$$\alpha \simeq \frac{f_{DA}}{f_{DD}} = \frac{g_{DA}}{g_{DD}} \tag{9.14}$$

Where f_{DA} and f_{DD} are the donor fluorescence detection rates (Sec.4.1.1) detected in the donor (DD) and the acceptor (DA) detection channels, respectively. In analogy, g_{DA} and g_{DD} are the total detection efficiencies of the donor emitted photons in the two detection channels. The working equation for the experimental determination of the cross-talk α is derived by expressing Eq.9.14 as a function of measureable quantities with Eq.4.8 ($f_f = \frac{s_f - bg_f}{N}$):

$$\alpha = \frac{s_{DA} - bg_{DA}}{s_{DD} - bg_{DD}} \tag{9.15}$$

Here, s_{DA} and s_{DD} are the total detection rates, and bg_{DA} and bg_{DD} the background detection rates in the two detection channels (D and A). The total count rates s_f were measured by FCS measurements of 2 min upon excitation of the single-labeled donor only samples (slDNA and slPGK). The background count rates bg_f were obtained from the same measurements performed on pure buffer. Afterwards, the cross-talks of the single-labeled samples were first calculated with Eq.9.15, and then the cross-talk of the double-labeled samples was estimated as described in Sec.9.2. The results are shown in Tab.9.4.

Table 9.4: Cross-talk α of double-labeled DNA and PGK.

sample	α
dlDNA	0.053
dlPGK	0.039

9.6 Förster Radius: R_0

The Förster radius R_0 is an important parameter, because it allows one to calculate the inter-dye distances from the donor lifetime $\tau_{DA}(r)$ (Eq.6.2) and from the energy transfer efficiency E(r) (Eq.8.2). In this thesis, Eq.8.3 was used to calculate R_0. The orientation factor was set to $k^2 = \frac{2}{3}$ because the dyes are freely rotating as was observed by steady state anisotropy measurements [32]. In addition, the refractive indeces n of the solutions were measured on the refractometer, and the donor quantum yields ϕ_D reported in Tab.9.2 were used together with the measured corrected fluorescence emission $f'_D(\lambda)$ spectra of the donor and with the normalized optical density spectra $OD'_A(\lambda)$ of the acceptor. The extinction spectra of the acceptor $\epsilon_A(\lambda)$ were calculated from the normalized optical density spectra $OD'_A(\lambda)$:

$$\epsilon_A(\lambda) = \epsilon_A(\lambda_{max}) \cdot OD'_A(\lambda) \tag{9.16}$$

Here, $\epsilon_A(\lambda_{max})$ is the extinction coefficient of the acceptor at the maximum peak position. For Al647: $\epsilon_A(\lambda_{max}) = 239000\,M^{-1}cm^{-1}$. The resulting Förster radii for DNA and PGK are

reported in Tab.9.5. Since, relevant differences were not observed, due to the different positions of the dyes on PGK or DNA, and due to the presence of the substrates, these details will not be discussed.

Table 9.5: Förster radii R_0 of the Al488/Al647 FRET pair bound to DNA and PGK. $\epsilon_A(\lambda_{ex}) = 239000\,M^{-1}cm^{-1}$, $k^2 = \frac{2}{3}$, $n = 1.334$.

sample	R_0 [Å]
dlDNA	51.1
dlPGK	51.0

9.7 Detection Efficiency Ratio: g'

The detection efficiency ratio $g' = \frac{g_A}{g_D}$ enters the calculation of the FRET estimators F' and E (Sec.8.3). Therefore, it must be measured in order to analyze the smFRET experiments. Two double-labeled DNA samples ("10bp" and "17bp") were used for the measurements of g'. In addition, a dye (Rh101) emitting in the acceptor and in the donor detection channel was used to correct for daily fluctuations of the detection efficiency g_f.

9.7.1 The Working Equation

The detection efficiency ratio g' was determined through calibration measurements by means of samples which do not show conformational dynamics. In fact, such behavior is described by the static-line (Sec.8.6), which means that the experimental populations on the 2D-plot must satisfy the following relation (Eq.8.12):

$$F' = \frac{1}{\gamma'} \cdot \frac{\tau_{DA}}{\langle \tau_{D0} \rangle_f - \tau_{DA}} \tag{9.17}$$

The equation shows, that once the quantum yield ratio γ' and the donor only intensity averaged lifetimes $\langle \tau_{D0} \rangle_f$ are known (Sec.9.3.3 and Sec.9.4), the expected correlation between the measured F' and τ_{DA} is easily calculated. Therefore, the detection efficiency ratio g' was derived by manually tuning its value until the experimental populations were superimposed on the expected behavior (see Fig.9.2). In fact, the vertical position of the population depends on $F' = \frac{F_D}{F_A} \cdot g'$ (Eq.8.5), that is calculated from the measured intensities (F_D and F_A) and from the unknown detection efficiency ratio g'.

9.7.2 The Calibration Measurement

In this work, two double-labeled DNA samples with different inter-dye distances ("10bp" and "17bp"), which correspond to different FRET efficiencies, were used to perform the calibration. DNA was choosen because its persistence length ($\ell_p \sim 500$ Å) [79] is ten times higher than the Förster radius ($R_0 \sim 50$ Å), meaning that on the distances sampled by FRET the DNA behaves like a rigid rod without internal dynamics. Therefore, only the fast motion of the dye linker must be considered. This contribution was included in the description by using the "corrected" static line (Eq.8.12). The whole measurements procedure is reported in Appx.D.6, together with the parameters used for the analysis. Here, only the results of the calibration are reported (Fig.9.2). A reference detection efficiency ratio $g' = 0.7$ was obtained.

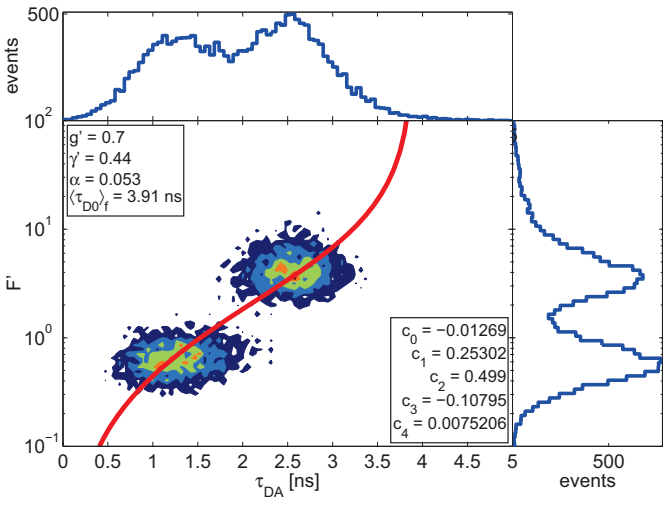


Figure 9.2: 2D-plot of the calibration measurements conducted with double-labeled DNA ("10bp" and "17bp"). The parameters used for the calculation of the "corrected" static line (red line) are shown in the insets together with the optimazed detection efficiency ratio g'. The following gaussian width was used to correct the static line: $\sigma_{DA} = 6$ Å (see Appx.D.6.2).

9.7.3 Daily Correction

As already mentioned before in Sec.4.1.1, the detection efficiency is made up of two contributions, the electronic detection efficiency g_{el} and the optical detection efficiency g_{opt}. The optical part g_{opt} is almost constant if the optical elements of the microscope are stable and unchanged in time ($g'_{opt}(\lambda) \simeq$ const). On the other hand, the electronic detection efficiencies

of the photodetectors g_{el} undergo small daily fluctuations. Therefore, to correct for the daily variations the following equation was used:

$$g'_U = g'_R \cdot \frac{\beta_U}{\beta_R} \tag{9.18}$$

Where, $g'_R = 0.7$ is given by the calibration measurement (Sec.9.7.2), and $\beta = \frac{f_A}{f_D}$ is the ratio of the measured fluorescence detection rates (f_A and f_D) of a dye emitting in both detection channels (A and D). Therefore, the unknown detection efficiency ratio g'_U was determined: (i) by measuring β_U, and (ii) by comparing it with the reference value $\beta_R = 31.82$ obtained the same day of the calibration measurement. Since Eq.9.18 is valid for a dye emitting in both detection channels, Rh101 was used for the daily correction. The following average value was obtained for the detection efficiency ratio: $\bar{g}'_U = 0.75 \pm 0.06$. For details about the measurement see Appx.D.6.3.

9.8 Summary

In this chapter, the parameters needed for the analysis of the smFRET data were obtained. A general expression, which allows one to measure the fluorescence quantum yield ϕ with a comparative method, was given (Sec.9.3.1). Three different methods were used; one of them was applied to cross-calibrate the reference quantum yields ϕ_R with the spectrofluorimeter and the absorption spectrometer (Sec.9.3.2), and two of them were employed to measure the unknown quantum yields ϕ_U of double-labeled DNA and PGK on the confocal microscope (Sec.9.3.3). From the donor and the acceptor quantum yields, the quantum yield ratio γ' was calculated (Tab.9.2). The donor only lifetime was measured (Sec.9.4), the cross-talk α was measured (Sec.9.5), and the Förster radius R_0 was also measured (Sec.9.6). Finally, a calibration measurement was performed with two double-labeled DNA samples in order to get the detection efficiency ratio g' (Sec.9.7). This value was corrected for daily fluctuations as described in Sec.9.7.3. The obtained parameters will be used for the analysis of the smFRET and the ensemble lifetime measurements performed on double-labeled PGK (Chap.10).

Chapter 10

PGK Interdomain Dynamics

10.1 Introduction

In this chapter, all the tools set up in the first part of the thesis are applied to study the static and dynamic properties of ligand-free and liganded phosphoglycerate kinase (PGK) from yeast. Here, the reader is taken on a journey towards the model refinement molded under the framework of the energy landscape, and lead by different techniques, each of them giving a glimpse of different PGK peculiarites. The ligand modulation of the conformation of the domains is first conveyed by ensemble lifetime FRET measurements (eTCSPC), and then confirmed by single-molecule FRET (smFRET) experiments; smFRET measurements which bring out brand new features about the dynamics, the flexibility, and the conformations of the PGK domains, and which permits one to make a statement about the substrate binding mechanism. Furthermore, FCS and the average burst duration, with their sensitivity to the protein spatial extension, reveal the complex intrinsic nature of the open and partially closed conformation of the domains observed by smFRET. Finally, an elastic network model of PGK is built up which allows one to envision the geometry of the catalytically relevant motions of the domains encoded in the topology of the protein structure.

10.2 The Substrates

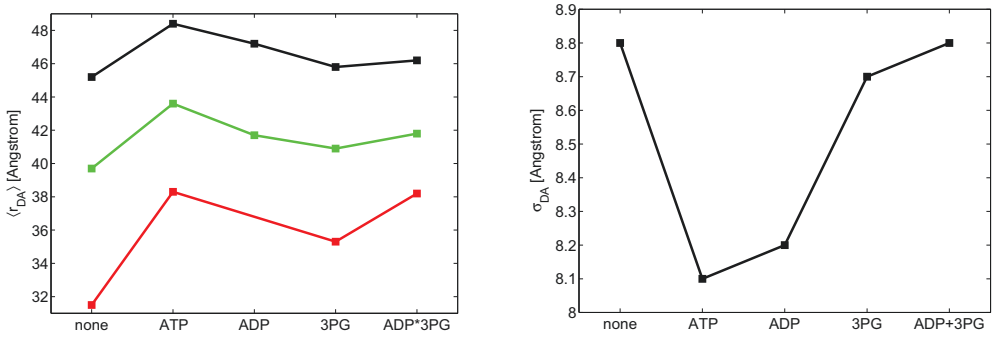
All the measurements done to study the binary and non-reactive ternary complexes of PGK (ATP, ADP, 3-PG and 3-PG*ADP) were performed with the double-labeled samples and a substrate concentration 10 times higher then the equilibrium dissociation constant (K_d), in order to have at least 90 % of bound PGK. The experimentally determined K_d values [32] are reported in Tab.10.1 together with the concentrations c_{sub} used for the measurements.

Table 10.1: Substrate equilibrium dissociation constants K_d and substrate concentrations c_{sub} used to perform the eFRET, the smFRET and the FCS measurements.

substrate	K_d [mM]	c_{sub} [mM]
Mg_2ATP	1.3	13
KAPD	/	13
3-PG	4.2	42

10.3 Ensemble Lifetime FRET Measurements

The ensemble lifetime FRET measurements were performed on ligand-free and liganded double-labeled PGK upon excitation of the donor as described in Sec.6.3. The donor intensity decays were fitted with a gaussian distribution of inter-dye distances $p_G(r)$ (Sec.6.3.2). The results are reported in Appx.E.2. Here, only the interesting information, namely the mean $\langle r_{DA} \rangle$ and the width σ_{DA} of the inter-dye distance distributions are discussed. The first parameter $\langle r_{DA} \rangle$ is related to the most probable conformation adopted by the protein domains. The second parameter σ_{DA} defines the amplitude of the fluctuations around the average inter-dye distance $\langle r_{DA} \rangle$ due to linker dynamics and/or motions of the protein domains (Eq.8.14).



(a) $\langle r_{DA} \rangle$ **(b)** σ_{DA}

Figure 10.1: Fitted parameters of the gaussian distribution of inter-dye distances $p_G(r_{DA})$ obtained from the fit of the donor intensity decays of ligand-free and liganded dlPGK. See Appx.E.2 for the full set of data. (a) Average inter-dye distances $\langle r_{DA} \rangle$: our data (black) and Haran et al. (red) [8]. Mean FRET averaged inter-dye distances $\langle r_{DA} \rangle_{E_1}$ (green). (b) Widths σ_{DA}.

In Fig.10.1a, the mean inter-dye distances $\langle r_{DA} \rangle$ measured with different ligands are shown (black) and compared with the data obtained by Haran et al. with the same methodology (red) [8]. The two data sets show the same qualitative behaviors, but with different absolute distances, with the substrate binding inducing small distance changes of $(2-20)\,\%$. Fur-

thermore, the same trend is observed by single-molecule FRET experiments (see Sec.10.4) shown in green for a conformation of the protein domains which resembles the partially closed crystal structure of PGK from yeast (pdb:1qpg). Therefore, assuming that the dyes do not interact with the substrates, these observations, which were obtained independently with different techniques, strongly indicate that the ligands fine tune the conformation preferentially adopted by the domains. However, the mean inter-dye distances $\langle r_{DA} \rangle$ measured in this work (black) at ensemble level seem to be biased towards higher values. In fact, values lower than the FRET averaged distances $\langle r_{DA} \rangle_E$ (green) are expected [33]. The most likely explanation is the magnitude of the donor only population ($\sim 0.57\,\%$), which precludes a precise estimation of the distance distributions [52]

Also the widths σ_{DA} are slightly changed upon substrates binding (Fig.10.1b). Here, a strong anticorrelation ($\rho = -0.926$) with respect to $\langle r_{DA} \rangle$ is observed. This would indicate that the ligands are also tuning the amplitudes of motions of the domains. However, the strong correlation between the fitting parameters (x_{D0}, $\langle r_{DA} \rangle$ and σ_{DA}) suggests that the real origin of the anticorrelation is the fitting procedure, even though a small contribution of the ligands cannot be excluded a priori. Therefore, a clear statement about the presence or the absence of motions of the domains in addition to the linker dynamics cannot be made, even though the observation of a distribution of distances implies that dynamics slower than the donor lifetimes ($\tau_c > \tau_{D0}$) are taking place [73].

More precise information about the dynamics timescales and the geometry of motions of the domains can be obtained by smFRET (Sec.10.4) and FCS (Sec.10.5) experiments, as well as by an elastic network model of PGK (Sec.10.6). In conclusion, it is worth to mention that the quality of the data and the estimated parameters can be improved by fitting the eTC-SPC histograms built up with all the photons selected by a burstwise analysis of the smFRET experiments, which allows one to discard the donor only species (see Sec.7.3).

10.4 Single-Molecule FRET Measurements

Single-molecule FRET measurements were performed on ligand-free and liganded double-labeled PGK. The results are discussed in this section, where the 1D efficiency histograms and the 2D-plots are the fundamental tools applied to the analysis and the interpretation of the single-molecule data. For more detailed information about the measuring procedure and the data analysis see Appx.E.3.

10.4.1 Ligand-Free PGK

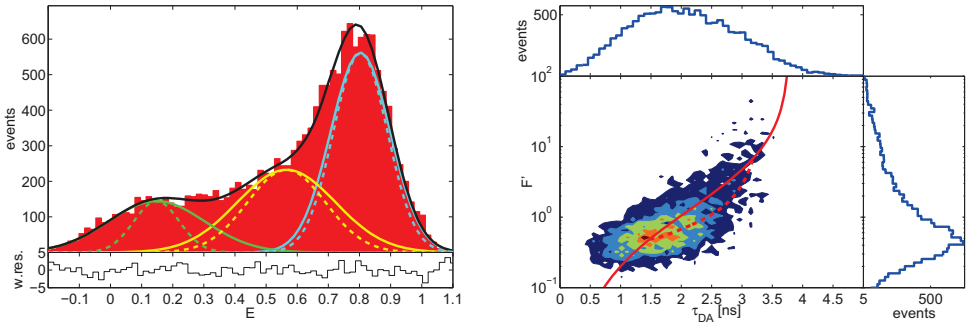
Ligand-free PGK is well described by a two states model where open and a partially closed conformations interconvert on the milliseconds timescale. This model is revealed by the 1D efficiency histogram (see Fig.10.2a), and by the 2D-plot (see Fig.10.2b), and gives new insights into the intrinsic dynamics of the PGK domains. Here, the 1D histogram was fitted with three gaussians which represent a partially closed conformation ($E_1 \sim 0.8$), an open conformation ($E_2 \sim 0.15$), and a mixed population ($E_{12} \sim 0.57$). The features of each population are discussed in this section.

The Partially Closed Conformation

The major population at $E_1 \sim 0.8$ (cyan continuous line) does not show conformational variance, which would appear as an extra width with respect to the dashed cyan line, and falls on the "corrected" static line at $\tau_{DA} \sim 1.5\,ns$ (continuous red line). These are clear evidences of dynamics faster than the observation time ($\tau_c << 1\,ms$) (Sec.8.6.3). Also here, as for the eFRET measurements (Sec.10.3), it is not possible to state if the motions of the domains contribute to the observed dynamics in addition to the linkers. Although, neutron spin echo spectroscopy (NSE) [11] strongly indicates that the protein contributes to the observed fast dynamics with motions of the domains at $\sim 60\,ns$. In addition, the mean FRET averaged distance of this population is $\langle r_{DA}\rangle_{E_1} = 39.7$ Å. This value is closer to the distance ($R_{135-290} = 39.3$ Å) between the C_α atoms of the labeling sites of PGK from yeast, which was calculated from the crystal structure of the protein in a partially closed conformation (pdb:1qpg) [4]. Therefore, the population at $E_1 \sim 0.8$ most likely represents a partially closed conformation of the domains with internal dynamics faster than milliseconds, which most probably has a characteristic time of a few tens of nanoseconds ($\tau_c \sim 60\,ns$). This means that, the energy profile of the system in proximity of the mimimum is a tight well with steep walls.

The Open Conformation

The population at $E_2 \sim 0.15$ (green continuous line), which appears in the 2D-plot at $\tau_{DA} \sim 3.2\,ns$, has conformational variance, as shown by the extra width with respect to the dashed green line, and is broadened along the "corrected" static line (continuous red line). Therefore, the underlying state distribution is static $p[r(t)] \sim p(r)$ on the milliseconds timescale ($\tau_c >> 1\,ms$) (Sec.8.6). The amplitude of the conformational space sampled by the protein domains is pretty large and is equal to 17.8 Å without correcting for the linkers contribution (see Appx.E.3.6), therefore revealing a flexible conformation sampling a wide energy profile. In addition, the relevant increase of the mean FRET averaged distances to $\langle r_{DA}\rangle_{E_2} = 67.2$ Å with respect to the partially closed state indicates an open conformation of the domains.



(a) 1D frequency histogram **(b)** 2D-plot

Figure 10.2: (a) 1D frequency histogram fitted with a 3-gaussian model function (Sec.8.5). The dashed lines represent the amplitude expected for a system without conformational variance σ_r^2 (see Sec.8.4). (b) 2D-plot with "corrected" static line (continuous line) and 2-state dynamic line (dashed line) connecting the partially closed ($E_1 \sim 0.8$) and the open ($E_2 \sim 0.15$) states. The set of parameters used for and retrieved from the analysis are reported in Appx.E.3.1.

The Mixed Population

The population at $E_{12} \sim 0.57$ (yellow line), which is centered at $\tau_{DA} \simeq 2.2\,ns$, shows conformational variance and is smeared along the 2-states dynamic line (dashed line) connecting the other two populations. Therefore, it indicates slow interconversion ($\tau_c \simeq 1\,ms$) between the highly flexible open conformation and the partially closed state. In fact, it represents those single-molecule events where the protein jumps between the low ($E_2 \sim 0.15$) and the high ($E_1 \sim 0.8$) efficiencies states during the observation time ($T \sim 1\,ms$). Finally, the fact that the mixed population at $E_{12} \sim 0.57$ is closer to $E_1 \sim 0.8$ is a consequence of the higher occupation probability of the partially closed state ($p_1 = 0.71$) with respect to the open one ($p_2 = 0.29$), and indicates that the partially closed state is energetically more stable and corresponds to a global mimimum.

Conclusions

This is the first time that the intrinsic dynamics of the PGK domains are detected with a direct measurement showing a dynamic equilibrium between an open and a partially closed conformation. Furthermore, the open conformation is rather loose and flexible, while the partially closed conformation is more rigid. In addition, the two conformations interconvert on the milliseconds timescale and have internal dynamics with characteristic times of a tens of nanoseconds (closed), and slower than milliseconds (open). These results, which were obtained

in the absence of ligands, highlight that PGK is an intrinsically flexible and dynamic system able to sample a wealth of conformations on largely different timescales $(10\,\text{ns} - 1\,\text{ms})$.

10.4.2 Liganded PGK

As shown in the left column of Fig.10.3, substrate binding has two major effects on the measured efficiency distributions in comparison with the ligand-free form of PGK (see Fig.10.2a). First, the open state ($E_2 \sim 0.15$) gets dramatically depopulated from $p_2 \sim 0.3$ to $p_2 \sim 0.1$ in favour of the partially closed one ($E_1 \sim 0.75$), which means that the relative stability of the partially closed conformation is increased. Second, the average position of the major population slightly decreases, as shown by the shift with respect to the vertical dashed blue line. This proves that a conformational change of the relative position of the domains, that increases the average inter-dye distance, takes place within the partially closed state. Here, the calculated mean FRET averaged distances $\langle r_{DA} \rangle_{E_1}$ (see Tab.E.5) show the same trend observed with the eFRET lifetime experiments (see Fig.10.1a). In addition, comparable absolute values are observed with respect to the data from Haran (red) considering the different averaging of the inter-dye distances and the different dye linkers (see Eq.E.25 and Eq.E.26). Therefore, this observation suggests that in the eTCSPC data, the partially closed population is mainly visible. To be more detailed, Mg_2ATP induces the largest conformational change of the domains with a relative change of $\sim 4\,\text{Å}$ ($\sim 12\,\%$) of the FRET averaged inter-dye distance. The other complexes present smaller relative changes of circa $(1-2)\,\text{Å}$ ($\sim (6-8)\,\%$).
Considering the dynamics of the system, the timescale of motions of the domains are not affected by the substrates in the time-resolution limits of the smFRET experiments. In fact, the mixed populations at $\tau_{DA} \sim 2.2\,\text{ns}$ are still smeared along the dynamic line, showing slow interconversion on the milliseconds timescale. In addition, the major populations ($\tau_{DA} \sim 1.5\,\text{ns}$) still lie on the "corrected" static line and do not show conformational variance, confirming the presence of dynamics faster than milliseconds. Nevertheless, a complementary technique, NSE, indicates that the dynamics of the domains become faster in the liganded state ($\tau_c \sim 30\,\text{ns}$) and the amplitude of motions is reduced [11]. Which means that the protein partially closed conformation becomes more rigid, and that the energy basin becomes steeper and narrower.

Conclusions

The smFRET experiments point out that the ligands adjust the conformation of the protein domains within the partially closed state inducing a state redistribution that increases the mean inter-dye distance. In addition, the ligands tune the equilibrium with the open conformation in favour of the partially closed state that is stabilized and gets more rigid.

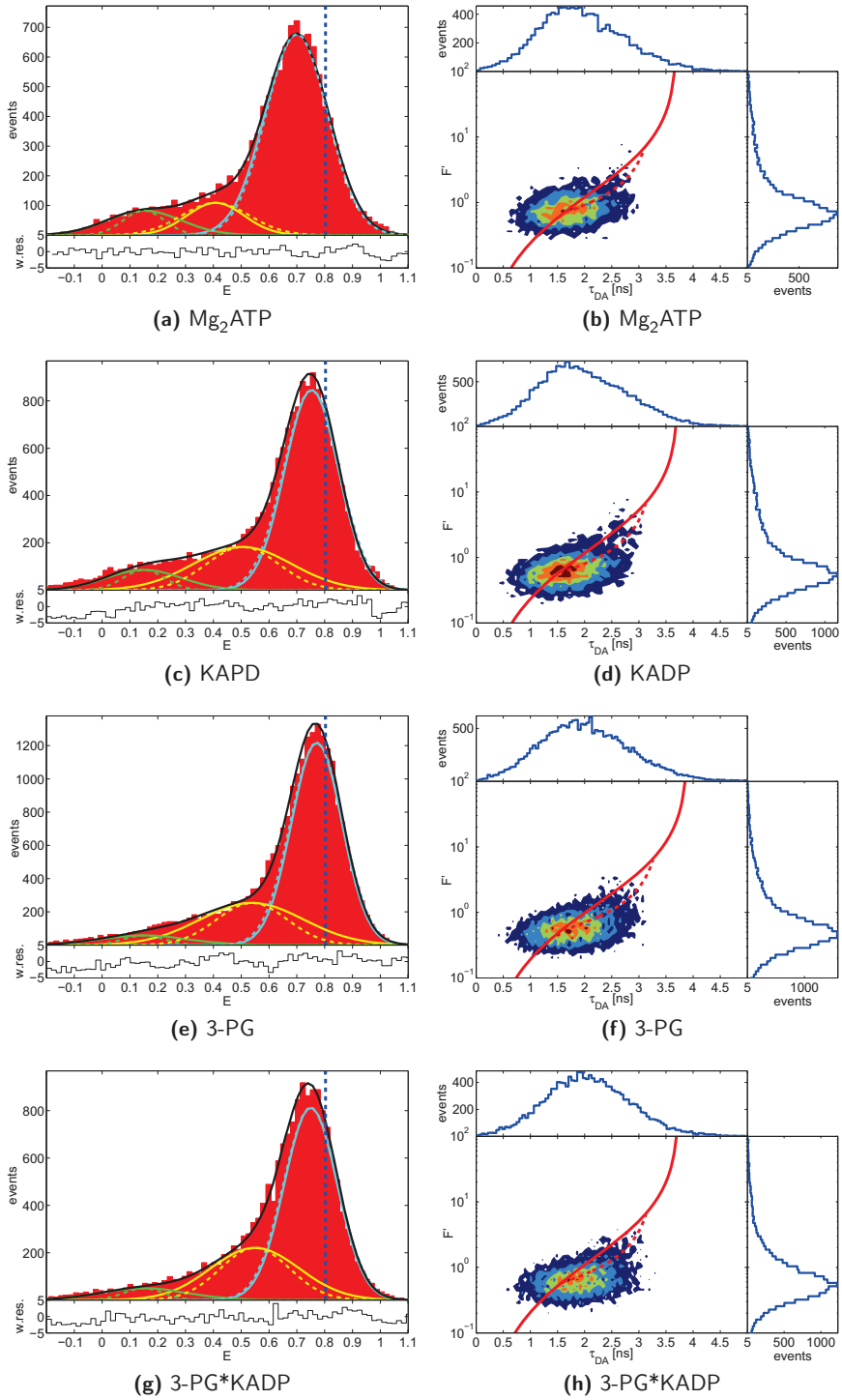


Figure 10.3: 1D efficiency histograms and 2D-plots of liganded PGK. The mean position of the low FRET population was fixed at $\langle E_2 \rangle = 0.15$ during the fitting procedure. The blue dashed vertical line indicates the position of the main population ($\langle E_1 \rangle = 0.80$) in the efficiency histogram of ligand-free PGK (Fig.10.2a).

10.4.3 The Energy Landscape of PGK

The energy landscape is a powerfull tool to describe the static and the dynamic properties of a protein at the thermodynamic equilibrium [14]. This concept was first introduced to explain the protein folding mechanism, and afterwards was extended to describe the interaction between a protein and a binding partner (ligands, effectors, macromolecules, etc.) [15]. Under this perspective, the conformational space sampled by the protein is represented by a multidimensional energy surface where individual points correspond to different conformations (Appx.E.3.7). The features of the energy profile depend on the intrinsic properties of the protein as well as on the external conditions (T, pH, salt, etc.). One of these effectors is the protein binding partner. Therefore, the binding mechanism between a protein and its ligand can be visualized by a dynamic energy landscape where the protein jumps between the energy profiles of its bound and unbound states [20]. Since these two profiles are generally different, the binding process may affect the static and the dynamic properties of the protein. Therefore, single-molecule techniques are particularly suited to study the characteristics of the energy landscapes and the binding mechanisms of a ligand to a protein. In fact, the conformational distribution of the system at equilibrium is directly recovered without ensemble averaging effects, and the time-range of the dynamics can also be estimated.
In this work, a glance at the energy landscape of PGK at the thermodynamic equilibrium is given by using the information recovered from the single-molecule FRET experiments, namely the state probability distributions and the dynamics timescales. Here, the energy landscape is depicted by a one dimensional energy profile $G(r_{DA})$ as a function of the inter-dye distance r_{DA} that is sensitive to the rigid body movements of the domains. The recovered state distribution allows one to calculate the equilibrium constants K_{ij} and the relative Gibbs free energies ΔG_{ij} between two sub-ensembles of states i and j from the occupation probabilities p_i and p_j (Appx.E.3.7 and Appx.E.3.8). Furthermore, the timescales and the amplitudes of motions of the domains in turn give an idea about the heights of the energy barriers $\Delta G_k^{\ddagger}$, and about the topology of the energy profile sampled by the protein. In fact, the dynamics of the system depends on the steepness and the height of the barriers, as well as on the roughness and curvature of the energy wells [15].

The Kinetic Scheme

The binding mechanisms of a ligand to PGK in the presence of the protein intrinsic dynamics between the open (PGK_O) and the partially closed (PGK_C) conformations of the domains can

be described by the following kinetic scheme:

$$\begin{array}{ccc} (PGK_O + S)_1 & \overset{K_{12}}{\rightleftharpoons} & (PGK_C + S)_2 \\ \Big\Updownarrow K_{14} & & \Big\Updownarrow K_{23} \\ (PGK_O \cdot S)_4 & \overset{K_{43}}{\rightleftharpoons} & (PGK_C \cdot S)_3 \end{array} \tag{10.1}$$

Here, the upper and the lower half of the scheme illustrate the protein interconversion between the open (O) and the partially closed (C) conformations in the ligand-free (PGK) and liganded ($PGK \cdot S$) forms, respectively. Under the energy landscape perspective, each of these states correspond to an ensemble of conformers with similar energies. Here, the extent of the ensembles depends on the topology of the energy profile in proximity of the minima. Finally, the vertical transitions represent the substrate binding and unbinding processes that can be visualized as jumps between the energy profiles of the ligand-free and liganded PGK.
As declared before, the relative Gibbs free energies ΔG_{ij} and the equilibrium constants K_{ij} between a pair of states can be obtained if the relative populations p_k of the four PGK species ($k = 1, 2, 3, 4$) appearing in the kinetic scheme are known. Unfortunately, at intermediate substrate concentrations ($c_{sub} \simeq K_d$), the liganded ($PGK_O \cdot S$ and $PGK_C \cdot S$) and the ligand-free (PGK_O and PGK_C) species coexist. As a consequence, the relative populations of the four states are not accessible because they are almost degenerate in efficiency. In fact, the open and the partially closed conformation have nearly the same transfer efficiencies with or without ligands (see Fig.10.2 and Fig.10.3). Luckily, the relative populations of ligand-free and liganded PGK can be quantified separately by pushing the equilibrium up and down by measuring at zero ($c_{sub} = 0$) and very high ($c_{sub} >> K_d$) substrate concentrations. In this way, the upper and the lower half of the kinetic scheme are evaluated independentely with a two states model (Appx.E.3.8), and the energy profiles $G(r_{DA})$ sampled by the protein in the bound and the unbound states can be reconstructed.

Topology of the Energy Landscape

The energy landscape of PGK can be envisioned with a tight energy basin separated by a fairly low energy barrier from a broad local mimimum (see Fig.10.4). Here, the tight well corresponds to a set of relatively similar partially closed conformations of the domains interconverting on the slow nanoseconds timescale. On the other hand, the broad energy mimimum corresponds to the open state, with the protein domains assuming a wide range of static conformations on the milliseconds time-regime. The main effects of the substrates on the energy profile are the narrowing and the small displacement of the absolute energy mimimum towards higher mean

FRET average distances ($\langle r_{DA,3}\rangle_{E_1} > \langle r_{DA,2}\rangle_{E_1}$ see Fig.10.1a), and the shift of the equilibrium towards the partially closed conformation ($p_3 > p_2$). However, drastic changes on the general topology of the energy profile are not observed upon addition of the ligands. In fact, the timescales and the amplitudes of the interdomain motions are not severely affected by the substrates (Sec.10.4.2).

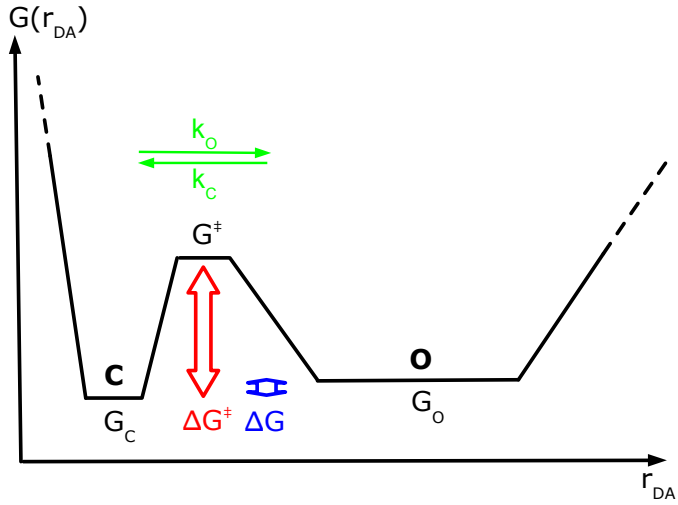


Figure 10.4: Sketch of the energy landscape $G(r_{DA})$ of PGK. The inter-dye distance r_{DA} has being used as the reaction coordinate. The two energy wells represents the partially closed ($C = 2, 3$) and the open ($O = 1, 4$) states shown in the kinetic scheme previously reported in Eq.10.1. The microscopic rate constants k_O and k_C represents the opening and the closing rates, respectively. $\Delta G^{\ddagger}$ is the activation energy and ΔG the relative Gibbs free energy difference between the two subensembles of states (C and O).

Quantitative Results

More quantitative results are given below, by computing the equilibrium constants (K_{12} and K_{43}) and the relative Gibbs free energies (ΔG_{12} and ΔG_{43}) for the kinetic scheme represented in Eq.10.1. This was done by using the occupation probabilities of the liganded (p_2 and p_3) and the ligand-free (p_1 and p_4) states reported in Tab.10.2. In addition, the activation energies ($\Delta G^{\ddagger}_{1/4}$ and $\Delta G^{\ddagger}_{2/3}$) for the closing and the opening transitions are also calculated. This is done by considering an approximated interconversion time τ_c between the open and the partially closed conformations, that is roughly estimated from the smFRET data.

The ligand-free state of PGK is first analyzed. Here, the relative free energies between the two mimina is $\Delta G_{12} = G_2 - G_1 = -0.53\,\frac{\text{kcal}}{\text{mol}}$, and the equilibrium constant is $K_{12} = 2.5$. Therefore, the partially closed state is more stable and the closing rate k_C is ~ 2.5 times larger than

the opening rate k_O; in fact $K_{12} = K_{43} = \frac{k_C}{k_O}$ (see Appx.E.3.8).
Considering the complexes of PGK with the substrates, the average effect of the ligands is to increase the energy difference between the open and the partially closed conformation of the domains to $\langle \Delta G_{43} \rangle = \langle -1.31 \rangle \frac{kcal}{mol}$. At the same time, the average equilibrium constant arises to $\langle K_{43} \rangle = \langle 9.5 \rangle$. Therefore, the ligands shift the equilibrium towards the partially closed conformation, increasing the closing rate k_C with respect to the opening rate k_O. To be more precise, the most stable complex is with 3-PG. Here, the energy difference $\Delta G_{43} = -1.52 \frac{kcal}{mol}$ is three times higher than in the ligand-free form, and $K_{43} = 13.3$. Consequentely, the closing rate is thirteen times higher than the opening rate ($k_C \simeq 13 \cdot k_O$). The binary complexes formed with the nucleotides moieties (ATP and APD) are less stable in comparison with 3-PG. In fact, the energy difference is $\Delta G_{43} = -1.17 \frac{kcal}{mol}$, and the equilibrium constant is $K_{43} = 7.3$, meaning that $k_C \simeq 7 \cdot k_O$. Finally, the stability of the ternary complex 3-PG*KAPD is intermediate between the binary complexes; in fact $\Delta G_{43} = -1.36 \frac{kcal}{mol}$ and $K_{43} = 10.1$.
The energy barriers $\Delta G^{\ddagger}_{1/4}$ overcome by PGK to jump from the open to the partially closed conformation, and vice versa $\Delta G^{\ddagger}_{2/3}$ were estimated considering that the characteristic time τ_c of the dynamics interconversion falls on the $(0.1 - 1)$ ms time regime (see Appx.E.3.9). The calculated values are reported in Tab.10.2.

Table 10.2: Occupation probabilities of the open (p_1 and p_4) and the partially closed (p_2 and p_3) state of the PGK domains in the ligand-free and liganded form obtained from the fit of the 1D efficiency histograms. The relative Gibbs free energies $\Delta G_{ij} = G_j - G_i$ and the equilibrium constants $K_{ij} = \frac{k_C}{k_O}$ were calculated from the occupation probabilities as described in Appx.E.3.7 and Appx.E.3.8. The activation barriers $\Delta G^{\ddagger}_i$ were estimated as described in App.E.3.9. All the parameters refer to Eq.10.1.

substrate	p_1	p_2	ΔG_{12} $\left[\frac{kcal}{mol}\right]$	K_{12}	$\Delta G^{\ddagger}_1$ $\left[\frac{kcal}{mol}\right]$	$\Delta G^{\ddagger}_2$ $\left[\frac{kcal}{mol}\right]$
none	0.29	0.71	-0.53	2.5	$(5.7 - 6.4)$	$(6.2 - 6.9)$
	p_4	p_3	ΔG_{43} $\left[\frac{kcal}{mol}\right]$	K_{43}	$\Delta G^{\ddagger}_4$ $\left[\frac{kcal}{mol}\right]$	$\Delta G^{\ddagger}_3$ $\left[\frac{kcal}{mol}\right]$
Mg_2ATP	0.12	0.88	-1.17	7.3	$(5.4 - 6.0)$	$(6.5 - 7.2)$
KADP	0.12	0.88	-1.17	7.3	$(5.4 - 6.0)$	$(6.5 - 7.2)$
3-PG	0.07	0.93	-1.52	13.3	$(5.2 - 5.9)$	$(6.7 - 7.4)$
3-PG*KAPD	0.09	0.91	-1.36	10.1	$(5.3 - 5.9)$	$(6.6 - 7.3)$
$\langle$substrate$\rangle$	$\langle 0.10 \rangle$	$\langle 0.90 \rangle$	$\langle -1.31 \rangle$	$\langle 9.5 \rangle$	$\langle (5.3 - 6.0) \rangle$	$\langle (6.6 - 7.3) \rangle$

Conclusions

The analysis of the experimentally determined conformational distributions of ligand-free and liganded PGK gives insight into the ligand binding mechanisms. In fact, the results summarized in the following scheme are consistent with the population-shift mechanism, also called "preexisting equilibrium/conformational selection" model [80][20][28]:

$$\begin{array}{ccc} (0.29)_1 & \underset{\rightleftharpoons}{\xrightarrow{2.5}} & (0.71)_2 \\ \Big\uparrow\Big\downarrow ? & & \Big\uparrow\Big\downarrow ? \\ \langle 0.10\rangle_4 & \underset{\rightleftharpoons}{\xrightarrow{\langle 9.5\rangle}} & \langle 0.90\rangle_3 \end{array} \tag{10.2}$$

Here, the occupation probabilities p_k of each species, and the equilibrium constants K_{ij} are shown for ligand-free (upper half) and liganded (lower half) PGK. The population-shift model postulates that the native state of a protein may not be defined by a single conformation, but rather by an ensemble of closely related conformations that coexist in equilibrium [80]. The most suitable conformers among these will bind the substrate, shifting the equilibrium towards the conformation best matching the ligand structural and dynamic properties. This is exactly what was observed for PGK. Here, an ensemble of partially closed and open conformations of the domains is sampled on different timescales by the protein in the ligand-free form and presented to the substrate that, once bound, shifts the equilibrium towards a preferential partially closed conformation (Eq.10.2). To be more precise, the ligands stabilize the partially closed conformation of the PGK domains by changing the energy imbalance between the activation energies ($\Delta G_C^{\ddagger}$ and $\Delta G_O^{\ddagger}$), that in turns leads to the modulation of the closing (k_C) and the opening (k_O) microscopic rates, as shown above.

Unfortunately, without additional information, it is not possible to determine what are the binding energies ΔG_b and the equilibrium association constants $K_a = \frac{[PGK{\cdot}S]}{[PGK][S]}$ for the two states. Therefore, the conformation which better lodges the substrates is not known. Only a series of measurements at different substrates concentrations may answer this question [81]. In addition, a measurement of the state occupation probabilities as a function of the temperature would increase the precision of the ΔG_{ij} values (see Eq.E.8). Furthermore, the precision of the estimated energy barriers $\Delta G^{\ddagger}$ can also be increased by measuring the characteristic time τ_c of the dynamics as a function of the temperature, and by making an Eyring plot. In this case, the entropy $\Delta S^{\ddagger}$ and the enthalpy $\Delta H^{\ddagger}$ of activation would also be obtained giving some hints about the possible microscopic mechanisms underlying the transition of the protein over the energy barrier. Finally, nothing can be stated about possible conformational rearrangements induced by the ligands internally to the domains. In this regards, intradomain FRET experiments could give some hints about this puzzle.

10.5 FCS Measurements and Average Burst Duration

The diffusion coefficients were determined by FCS experiments with and without substrates, as described in Sec.5.5. The calculated hydrodynamic R_h (Eq.5.9) and gyration R_g (Eq.5.10) radii are reported in Tab.10.3. In Fig.10.5, the radii of gyration measured with FCS are compared with the values obtained by small-angle diffraction experiments (SAXS) on PGK from yeast and human.

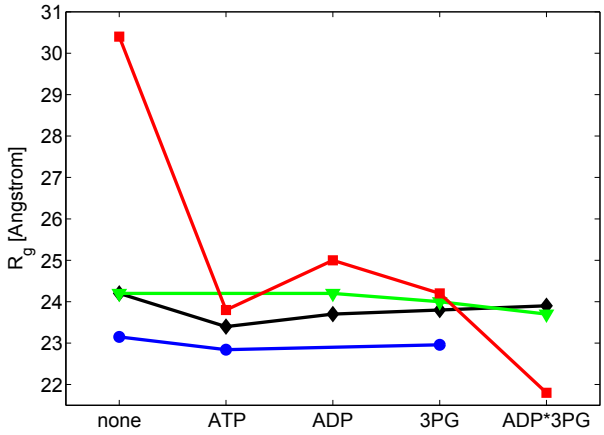


Figure 10.5: Radii of gyration R_g of ligand-free and liganded PGK meausured by FCS (red) and SAXS (other colors). The SAXS measurements were performed on human [7][6] (black and blue) and yeast PGK [82] (green).

By looking at the FCS data in red, the most striking effect prompted by the substrates is the decrease of the radius of gyration by $\sim 20\,\%$ with respect to ligand-free PGK. Since the data were corrected for the different viscosities of the solutions (Appx.E.4) as well as for the change in the refractive indeces (Appx.E.5); and furthermore as the temperature was kept constant, the only explanation is the decrease of the radius of gyration R_g.
This behavior is explained assuming that the open state, which was observed at low efficiency ($E_2 \sim 0.15$) (see Sec.10.4), corresponds to a partially expanded conformation of the protein (high R_g) while the partially closed state corresponds to a compact conformation (low R_g). In fact, under this assumption and considering equal visibility of the open and the partially closed species in the FCS data ($f_O \simeq f_C$) [84][42], even though a bias towards the open state is more feasible, the measured diffusion coefficient of ligand-free PGK would be an ensemble averaged value $\langle D \rangle = 59.2\,\frac{\mu m^2}{s}$ between the diffusion coefficients of the two conformations (D_O and D_C) weighted by their occupation probabilities (p_O and p_C):

$$\langle D \rangle = p_O \cdot D_O + p_C \cdot D_C \tag{10.3}$$

Table 10.3: The diffusion coefficients D, the viscosities η and the refractive indexes were measured at 23 °C. The measured diffusion coefficients were corrected for the refractive index mismatch of the buffer solution with respect to water ($n_{H_2O} = 1.333$), by using the dependence between the apparent diffusion coefficient and the refractive index calculated by Enderlein at al. [83]. The radii R_g and R_h were calculated with Eq.5.10 and Eq.5.9, respectively.

substrate	$D^{23\,°C}$ $\left[\frac{\mu m^2}{s}\right]$	η [mPa · s]	n	$D_{corr}^{23\,°C}$ $\left[\frac{\mu m^2}{s}\right]$	R_h [Å]	R_g [Å]
none	58	0.950	1.3340	59.2	38.6	30.4
Mg_2ATP	68	0.973	1.3353	73.9	30.2	23.8
KADP	65	0.970	1.3355	70.7	31.6	25.0
3-PG	66	0.988	1.3354	71.7	30.6	24.2
3-PG*KADP	66	1.010	1.3369	77.6	27.7	21.8

Since, in the liganded form the occupation probability of the open state is low compared to the partially closed state ($p_C \geq 10 \cdot p_O$) (see Tab.10.2), Eq.10.3 simplifies to $\langle D \rangle \simeq D_C$. Thus, the diffusion coefficient of the partially closed state is well approximated by the diffusion coefficient measured for the liganded form and averaged over the different substrates conditions $D_C = \langle 73.5 \rangle \frac{\mu m^2}{s}$ (see Tab.10.3). This allows one to calculate the diffusion coefficient corresponding to the open conformation with Eq.10.3 by using the occupation probabilities ($p_C = 0.71$ and $p_O = 0.29$) of the ligand-free form of PGK (see Eq.10.2):

$$D_O \simeq \frac{\langle D \rangle - p_C \cdot D_C}{p_O} \tag{10.4}$$

This gives an underestimated value of $D_O \sim 23 \frac{\mu m^2}{s}$. This low value can be explained by small oligomers with an average approximated radius of gyration of ~ 75 Å, therefore showing the propensity of the PGK open conformations to interact with each other. Furthermore, this observation suggests that the open state detected by smFRET at $E_2 \sim 0.15$ is an ensemble of locally unfolded and extended conformations in dynamic equilibrium with each other, and with the oligomers. Here, the oligomerization is most probably driven by the hydrophobic interactions between locally unfolded conformations, which tend to bury the exposed hydrophobic surfaces to lower their Gibbs free energy [81]. Now the question is, why do the oligomers disappear when the protein is in the liganded form, as shown by the drastic decrease of R_g? The easiest explanation is the depletion of the open protein conformation upon substrate binding, which shifts the equilibrium towards the partially closed state and reduces the oligomerization probability. The following kinetic scheme is proposed to describe this process:

$$\{A_N\} \rightleftharpoons \{PGK_i\}_O \rightleftharpoons \{PGK_C\} \tag{10.5}$$

Where, $\{A_N\}$ is a set of oligomers built up by a variable number of monomers N, $\{PGK_i\}_O$ is a set of open, extended, flexible and locally unfolded conformers, and $\{PGK_C\}$ is a smaller ensemble of compact, more rigid and partially closed conformations. Here, the dynamic of the open conformers $\{PGK_i\}_O$ is slower than milliseconds, the partially closed conformations $\{PGK_i\}_C$ interconvert on the nanoseconds timescale, and the transitions between the open and the partially closed ensembles of states ($\{PGK_i\}_O \rightleftharpoons \{PGK_i\}_C$) take place on the milliseconds time-interval.

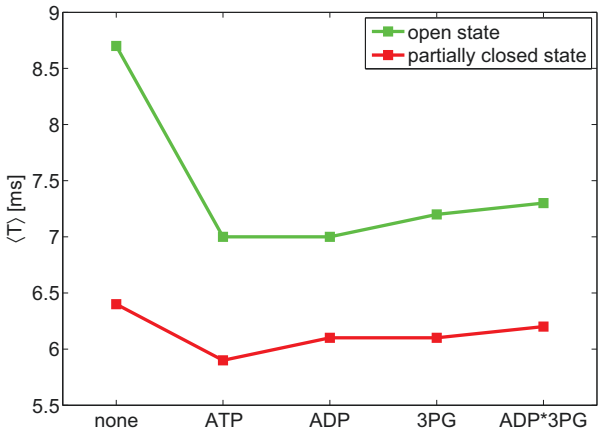


Figure 10.6: Average burst duration $\langle T \rangle$ of the open (green) and the partially closed (red) FRET populations. Bursts with $E \lesssim 0.2$ (open) and $E \gtrsim 0.8$ (partially closed) where choosen in order to discard the events belonging to the mixed population (see Fig.10.2a). More details are found in Appx.E.6.

The same picture is confirmed by the average duration $\langle T \rangle$ of the fluorescence bursts belonging to the open ($E_2 \sim 0.15$) and the partially closed ($E_1 \sim 0.8$) FRET populations (see Fig.10.6). In fact, the burst duration $\langle T \rangle$ allows one to take a closer look at the individual diffusive properties of the two sub-ensembles of conformations, and at the spatial extension of the partially closed and the open protein states. As shown in Fig.10.6, the average burst duration within the open state (green), which in the context of the previous model (Eq.10.5) is described by the equilibrium between the oligomers and the open conformations of the PGK domains ($\{A_N\} \rightleftharpoons \{PGK_i\}_O$), drastically decreases upon addition of the substrates. Thence, something big, like the oligomers $\{A_N\}$, is diffusing, which decreases its population upon ligand binding, therefore corroborating the model previously proposed (Eq.10.5). A carefull analysis of the average burst duration $\langle T \rangle$ of the partially close state (red) allows also to envision the geometry of interdomain motions tuned by the ligands. In fact, the mean burst duration of the partially closed state slightly decreases upon substrate binding, thus supporting the SANS and SAXS

results [11][6][7] showing a more compact conformation of the liganded PGK. In addition, $\langle T \rangle$ is anticorrelated ($\rho = -0.905$) with respect to the mean FRET averaged distance $\langle r_{DA} \rangle_{E_1}$ (see Fig.10.1a). Therefore, the ligands shift the equilibrium towards a more compact conformation of the domains having higher average inter-dye distance with respect to the partially closed conformation adopted by PGK in the ligand-free form. To conclude, the transition tuned by the ligands takes place along a preferential trajectory which displays an anticorrelated behavior between the inter-dye distance ($\langle r_{DA} \rangle_E$) and the protein spatial extension (R_g).
Under the same perspective of Eq.10.5, the big difference observed in the ligand-free form between the FCS and the SAXS data (see Fig.10.5) is explained by the depopulation of the open state induced by crowding effects. In fact, the SAXS measurements were performed at high protein concentrations ($> 100\,\mu$M), where protein-protein nonspecific interactions were shown to stabilize more compact conformations of the PGK domains [85], while the smFRET and the FCS experiments were conducted at ~ 10 pM and ~ 1 nM, respectively.

Conclusions

The study of the diffusive properties of PGK open up a more complex scenario with respect to what was observed by smFRET experiments (see Sec.10.4). In particular, a closer look at the open state, which was observed at low efficiency $E_2 \sim 0.15$, brings out a complex intrinsic nature of this state that can be better described by an ensemble of interconverting oligomers $\{A_N\}$, and locally unfolded open conformations $\{PGK_i\}_O$. This observation is not surprising, because at the final end, the native energy landscape corresponds to the bottom of the folding funnel [15]. In fact, a similar population was already observed during unfolding studies [86]. Therefore, the transition between the open and the partially closed states may well be a heritage of the last step of folding. The hinge region is the best candidate to explain the locally unfolded conformations of PGK. In fact, this region is flexible, and a loss of its native structure would explain the falling apart of the domains inducing the observed increase of the inter-dye distance, and explaining the broad range of observed conformations. In addition, it is feasible that the hinge is the last structural element to fold, after the folding of the domains. Furthermore, the hinge can be destabilized by the histidine tag added at the N-terminus of the amminoacidic sequence, in order to fish the protein during the purification procedure [32]. In fact, the N-terminus is unstructured and binds to the surface of the hinge region. Here, a better solution would be to cleave the histidine tag. Additional measurements are nevertheless required to confirm this scenario, and in order to envision a possible microscopic mechanism for the unfolding of the hinge and for the crossing of the activation barriers, which separate the open and the partially closed states. In this regard, smFRET experiments under unfolding and crowding conditions would give insight into the relative stability of the two states (ΔG_{ij}), as

well as the energetics of the transition state ($\Delta G^{\ddagger}$, $\Delta H^{\ddagger}$ and $\Delta S^{\ddagger}$). Furthermore, the diffusion coefficients of the open and the partially closed conformations can be better assessed through sub-ensemble FCS measurements. Finally, the question is raised about the functional relevance of the open conformation, which can be investigated by changing the substrate concentrations in order to estimate the binding affinities $\frac{1}{K_a}$ for the open and partially closed states.

10.6 Normal Mode Analysis (NMA) and a Coarse-Grained Simulation

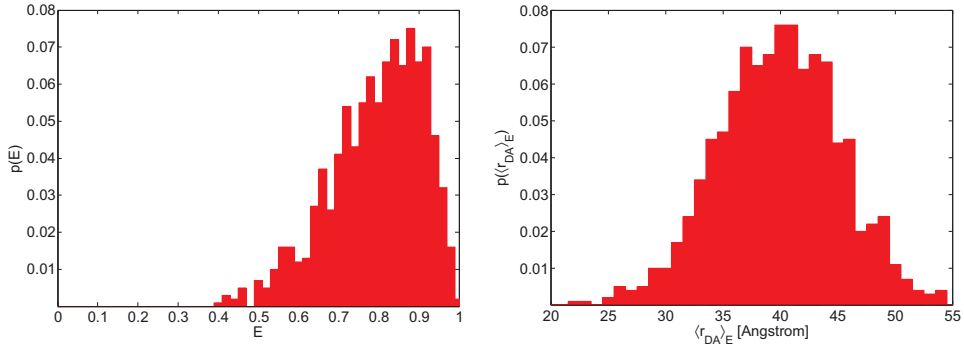
Normal modes analysis (NMA) is a powerful tool to predict the direction of the global motions of proteins. A series of studies based on this approach showed that the global dynamics of proteins are dictated by the overall architecture or topology of the interresidue contacts, and are insensitive to the structural details [18]. Furthermore, the intrinsic protein dynamics offer a few well defined energetically favorable mechanisms of structural rearrangements along preferential modes, and the ligand selects the one that best matches its structural and dynamic properties. The only condition for the application of the method is the knowledge of a reasonable crystal structure.

The Model

In this work, an elastic network of PGK was simulated under the random Multiparticle Collision Dynamics (rMPC) approach [31] (in collaboration with Simoń Poblete, ICS-2, Forschungszentrum Jülich), in order to give a more detailed picture of the domain motions within the PGK closed state in thermal equilibrium, and the relation of these motions to ligand binding and function. The constant of the harmonic bonds of the force field were scaled globally in order to roughly match the amplitude of the fluctuations of the distance between the labeled α-carbons (Q135 and S290) measured by Haran et al. ($\sigma_{CC} = 7.6$ Å) [8] within $\sim 10\,\%$. In this way, the model is independent from the measurements presented in this thesis. Finally, the dye linker contributions were calculated along the trajectories with the accessible volume alghoritm (AV) [33]. The whole method is described in Appx.E.7.

Comparison with the Experimental Data

In Fig.10.7 are displayed the probability distributions of the FRET efficiency p(E) and the FRET average inter-dye distances $p(\langle r_{DA}\rangle_E)$, calculated from the simulated trajectories. The ensemble averaged values calculated from the two probability distributions $\langle r_{DA}\rangle_E = 39.9$ Å and $\langle E\rangle = 0.8$ are in excellent agreement with the measured values of the partially closed population



(a) Transfer efficiency probability distribution: p(E) **(b)** FRET averaged inter-dye distance probability distribution: $p(\langle r_{DA}\rangle_E)$

Figure 10.7: Probability distributions of the transfer efficiency p(E) and the FRET averaged inter-dye distances $p(\langle r_{DA}\rangle_E)$ obtained by histogramming the calculated trajectories. See Appx.E.7.4. Data provided by Simoń Poblete.

of ligand-free PGK, that are $\langle r_{DA}\rangle_E = 39.7\,\text{Å}$ and $\langle E\rangle = 0.8$ (see Tab.10.4). Therefore, the internal dynamics of the partially closed conformation of ligand-free PGK can be analyzed in detail with the present model. Furthermore, the disagreement between the measured and the calculated $\langle r_{DA}\rangle$ values confirms, that the distance distributions obtained from the eFRET experiments are biased towards longer distances (see Sec.10.3).

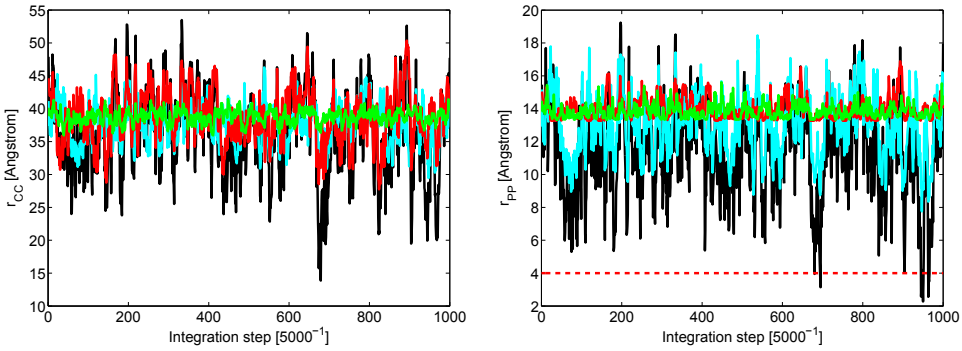
Table 10.4: Measured and calculated, ensemble ($\langle r_{DA}\rangle$) and single-molecule ($\langle r_{DA}\rangle_E$) averaged inter-dye distances and transfer efficiencies ($\langle E\rangle$). The simulated data are provided by Simoń Poblete.

approach	$\langle r_{DA}\rangle$ [Å]	$\langle r_{DA}\rangle_E$ [Å]	$\langle E\rangle$
e-/smFRET	45.2	39.7	0.80
rMPC	37.4	39.9	0.80

The Character of the Large Amplitude Interdomain Motions

The fluctuation of the distance r_{CC} between the labeled C_α (Q135 and S290) and the three slowest normal modes were calculated and used to characterize the movements of the domains. The trajectory reported in Fig.10.8a (black) shows that the C_α-distance fluctuates around an average value of 36.6 Å, with a FWHM of 15.3 Å. The maximum excursion of the motions reaches ~ 40 Å. Therefore, the domains perform large amplitude movements. The geometry of these motions is described by the three slowest normal modes: the hinge bending motion, the rocking motion, and the propelled twist motion. The hinge bending is a "Packman-like"

motion that closes the cleft between the domains (see Fig.1.1). The rocking motion is a rotation of the domains around the axis which is perpendicular to the cleft and to the axis crossing the two domains. The propelled twist motion is a torsion around the axis intersecting the domains with the N-terminal and the C-terminal domain moving against each other. The modes which contribute most to the large amplitude motions of the domains are the hinge bending (cyan) and the propelled twist (red).



(a) r_{CC} **(b)** r_{PP}

Figure 10.8: Trajectory of (a) the r_{CC} distance and the (b) r_{PP} distance (black), and their projections on the normal modes: hinge bending (cyan), propelled twist (red) and rocking (green). Data provided by Simoń Poblete.

Functional Motions of Ligand-Free PGK

The presence of possible functional movements in ligand-free PGK was assessed by looking at the distance r_{PP} between the 3-PG phosphate group and the γ-phosphate of the ATP-analog AMP-PNP, resolved by the crystal structure (pdb:1qpg). The distance of ~ 4 Å between the transferring phosphate in the PGK fully closed catalytically competent conformation [5] was used to determine the presence of enzymatic active conformations. The trajectory reported in Fig.10.8b (black line) shows that in average, the phosphates are too far to react, but once in a while thermal fluctuations drive a transition towards the fully closed catalytically active conformations (red dashed line). These transitions take place with a hinge-bending like motion (cyan line). Conversely, the other two modes (red and green lines) do not contribute to bring the substrate together. Therefore, the path towards the catalytically relevant conformation is encoded within the PGK topology which dictates the character of the large amplitude functional motions.

A Propelled Twist Motion Generated by the Substrates

The calculated FRET averaged inter-dye distance $\langle r_{DA} \rangle_E$ and the radius of gyration R_g are represented in Fig.10.9 for arbitrary displacement along the three slowest modes. The mutual dependences between the two quantities were used to describe the anticorrelated behavior observed upon substrate binding with smFRET experiments (see Sec.10.5). In fact, the only motion able to describe the anticorrelation between $\langle r_{DA} \rangle_E$ and R_g is the propelled twist motion shown in red (see Fig.10.9). Therefore, the substrate induced conformational rearrangement takes place along a trajectory described by the propelled twist motion, and is encoded in the 3D protein structure.

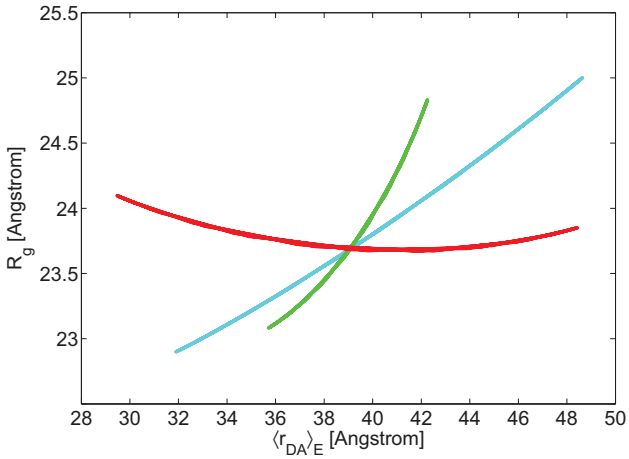


Figure 10.9: Values of R_g and $\langle r_{DA} \rangle_E$ for animated motions along the three slowest modes: hinge bending (cyan), propelled twist (red) and rocking (green). Data provided by Simoń Poblete.

Conclusions

The elastic network model of PGK simulated with a rMPC approach well describes the intrinsic dynamics of the partially closed state of the domains observed with smFRET experiments. The model shows that the enzyme undergoes big amplitude interdomain movements, and that the path towards the functionally relevant PGK conformation is encoded within the structure of the protein. Here, thermal fluctuations induce bursts of closure with a hinge bending like motion, which brings the substrate to the enzymatically active conformation. In addition, substrate binding induces a structural rearragement along a trajectory well described by a propelled twist motion, which as well as the hinge-bending depends on the protein topology.

Chapter 11

Conclusions

The Facts

In this thesis, the methodology to perform reliable and reproducible single-molecule FRET (sm-FRET) experiments at Angstrom resolution was established following the footsteps of Seidel and co-workers, Enderlein and co-workers, and Gopich and co-workers. The whole methodology was succesfully applied to investigate the interdomain dynamics of ligand-free and liganded phosphoglycerate kinase (PGK) from yeast. A model of the protein dynamics was constructed with an elastic network and simulated under a random Multiparticle Collision Dynamics (MPC) approach. The enzyme revealed itself as a flexible and dynamic system, with intrinsic dynamics on the nanosecond to millisecond timescales. A partially closed and compact conformation of the domains performing large amplitude motions faster than milliseconds was disclosed. A set of open and extended conformations of the domains in slow equilibrium, and with the tendency to oligomerize was observed. This state most probably corresponds to a late step of folding. The open and the partially closed states interconvert on the millisecond timescale. Addition of the ligands depopulates the open conformation and triggers a conformational redistribution within the partially closed state along a trajectory described by a propelled twist like motion. The path towards the enzymatically active conformations is encoded in the topology of the protein structure, which dictates the character of the big amplitude interdomain motions. Here, burstlike hinge bending transitions towards the reactive fully closed state of PGK are driven by thermal fluctuations.

The Impact of the Work on the PGK Studies

In the context of the PGK studies, the information obtained by the present work complements and confirms the results in the literature, extending the comprehension of the enzyme behavior in solution. In addition, an explanation for the apparently contradictory ensemble FRET data [8], and the small angle diffraction experiments [7][6][82][11] is given. In fact, the

anticorrelated modification of the inter-dye distance and the radius of gyration upon binding of the substrates is explained here by a conformational rearrangement, which takes place along a propelled twist like trajectory. Furthermore, a new state, the fully extended open state, is detected, and a feasible ligand binding mechanism is proposed. Moreover, for the first time, the interconversion between an open and a partially closed conformation of the domains is monitored with direct evidences. In fact, never before have the state distributions of PGK been investigated at single-molecule level.

The results presented here supplement previous works (see Chap.1), leading to the following description of PGK in solution. Ligand-free PGK in solution interconverts on the milliseconds timescale between an open and a partially closed conformation of the domains. The partially closed conformation has internal dynamics with a characteristic time of a few tens of nanoseconds [11][9], which take place around a set of hinge points [9] by way of mainly propelled twist and hinge bending like motions. From this partially closed state, hinge bending bursts of closure, which are thermally driven, bring the protein to the fully closed catalytically competent state. In the liganded form of PGK, the electrostatic interactions between the reactive substrates stabilize the fully closed conformation for the extremely short time of the phosphate transfer reaction [82]. Afterwards, the substrates are released by a spring loaded mechanism, which pulls the ligands apart [82]. Furthermore, ligand binding depopulates the open state and induces a conformational rearrangement within the partially closed state, taking place with a propelled twist like motion encoded in the topology of the protein structure. In addition, the substrates augment the protein rigidity [11][9][8], speed up the fluctuations of the domains [11][9], and increase the directionality of the motions by a reduction of the hinge points [9]. Finally, crowding effects, similarly to the ligands, depopulate the open conformation [85].

An Hypothetical Model of PGK's Reaction Cycle

With the present description, the following reaction mechanism is proposed for the enzyme. PGK in solution randomly fluctuates, presenting to the ligands a set of open and partially closed conformations, which sometimes reach the catalytically active state. The substrates can bind to each of these conformations with different binding affinities, not yet determined, and can select the ones best fitting their properties. The protein responds to substrate binding shifting the conformational equilibrium towards the partially closed state by means of a propelled twist like motion. This movement most likely positions the ligand in the proper geometry for the phosphate transfer reaction to occure. In addition, the substrate locks the protein, admitting mainly hinge bending like transitions towards the active conformation. Here, the ligand must await for the partner and hold on, until a causal stroke of the domains transport them on the catalytically compentent state. If this happens, the reaction take place and the subsequent

product release driven by a hydrophobic spring terminates the reaction cycle; otherwise, the ligands abandon the binding sites without reacting. Unfortunately, the on and off rates of the substrates, and the rate of occurrance of the transitions towards the enzymatically relevant conformation are not yet known. Therefore, it is still not possible to determine the rate limiting step of the full catalytic cycle.

The Impact of the Work on the Modern View on Protein Dynamics

The results presented above support the modern view on protein functional dynamics and molecular recognition [87][15][16]. In fact, phosphoglycerate kinase has a dynamic energy landscape [20] encoded in the 3D structure of the enzyme [27][19], where the intrinsic protein motions [19][18] occur preferentially in the direction of the catalytically competent fully closed conformation of the domains [23]. Furthermore, the ligands dynamically reshape the distribution of states restricting the range of conformational fluctuations, and shift the equilibrium by increasing the closing rate as already observed for adenylate kinase [27]. Thus, the evolution of the energy landscape by natural selection of the substates and pathways needed for function [23][19] appears to be a general feature of multidomain enzymes. Therefore, the emerging paradigm that *'structure-encodes-dynamics'* and *'dynamics-encodes-function'* [19] is confirmed once more.

Open Questions About PGK

What is still missing to give a complete description of the PGK catalytic cycle? In this thesis, new insights into the behavior of PGK in solution were presented. Nonetheless, a host of questions are not yet answered, and new problems arise. Above all, the nature of the open state is still not fully understood. Is this state artificially induced by the genetic modification of the protein through the histidine tag, or is this sub-ensemble of conformations functionally relevant? Can some hints about the last step of folding be inferred from a detailed investigation of this state? Even if the open state is not artificially created during the protein manipulation, what happens in a crowded cell environment? And the ligands... Do they prefere to bind to the open conformation, or to the partially closed one? Is the solvent flux generated by the interdomain fluctuations involved in the expulsion of the products, or capture of the reagents? And what is the bottleneck of the reaction cycle? Is it the probability to find both substrates on their binding positions, or the casual strokes of the domains towards the reactive conformation? All these questions and even more can be adressed by further experiments and different techniques.

Outlook

In the present work, the methodological premises to perform reliable and precise smFRET experiments were established. These achievements disclose a full set of experiments and further improvements, that can be performed on the short, the medium, and the long period of time. A list of the possible lines of investigation is given here:

1. **Sample:** New interdomain mutants of PGK are needed to increase the available information about the motions of the domains. In fact, more reference points would allow one to perform a triangulation to recover the geometry of the rigid motions of the domains, as well as permit one to refine the elastic network model. Furthermore, labeling sites with an ortogonal chemistry would strongly simplify the labeling process, as well as the measuring procedure and the data analysis. The approach recently developed by Lemke and co-workers [88] is the method of choice to replace one of the two cysteine labeling sites. In addition, the cleavage of the histidine tag is strongly requested to rule out possible effects on the hinge stability. Finally, the test [89][90] to determine the enzymatic activity must be improved to give quantitative and comparable results. The test must also be performed on the labeled protein in order to verify the effect of the dyes on the enzyme functionality.

2. **Exploration of the Parameter Space:** A wealth of experimental parameters can be modified to extract information about the dynamics of the domains, the ligand binding mechanisms, the folding process, the energetics and the functional behavior of PGK in solution.

    - Substrates Concentration: The variation of the substrates concentrations would allows one to determine the binding constants and energies [81][91][92] of the open and the partially closed conformations of the domains. This would give more hints about the functional relevance of the open state, and about the ligand binding mechanism.
    - Denaturants: Measurements at mild denaturant conditions would allow one to test if the transition between the open and the partially closed conformation of the domains is a vestige of the last step of folding.
    - Crowders: The use of crowding agents mimicking the cell environment would allow one to test if protein compaction correlates with the enzymatic activity as already observed for other proteins [93].
    - pH: The effect of pH is very prominent in enzyme kinetics. Furthermore, it was shown that under glucose starvation and sorbic acid stress the pH within the cell

compartments of *Saccharomyces cerevisiae* changes [94]. Therefore, it would be interesting to study how the motions of the PGK domains react to different physiological pHs, and possibly if the enzymatic activity is correlated with the eventual changes. In fact, PGK is a key enzyme of the glycolitic pathway.

- Temperature: Measurements at different temperatures would allow one to better determine the Gibbs free energies between the open and the partially closed state from the occupation probabilities of the two conformations (see Eq.E.8). Moreover, measuring the characteristic time of the dynamics at different temperatures would permit one to precisely evaluate the activation barriers, as well as the entropic and enthalpic contributions by means of an Eyring plot.
- Viscosity: Intracellular viscosities could slow protein conformational changes significantly and therefore influence the kinetics of the protein function [95]. Therefore, it would be interesting to study how the timescales of the interdomain dynamics of PGK changes at physiological viscosity $(2\,\mathrm{cP} - 3\,\mathrm{cP})$, and if this could change the rate limiting step of the phosphotransfer reaction.

3. **The Fully Closed Conformation:** The fully closed conformation of the PGK domains must be studied by stabilizing the transition state analogue complex (ADP*AlF$_4^-$*3PG) by means of aluminium tetrafluoride [82].

4. **Data Analysis Tools:** A lot of information is still hidden in the smFRET data. To unveil this information, the data analysis of the efficiency histograms must go beyond the empirical fit performed in this work. For instance, the rate constants of the slow interconversion between the open and the partially closed state of PGK can be evaluated more precisely by means of the photon distribution analysis (PDA) [70], by decoding the pattern of photons colors [96][97], by a model based fitting procedure of the efficiency histograms [69][97], or by the burst variance analysis (BVA) [98]. Furthermore, the single burst lifetime values can also be improved by a fit with a maximum likelihood estimator [55]. Finally, better ensemble inter-dye distance distributions can be obtained from the ensemble TCSPC histograms built up with the photons collected from the set of bursts of a single-molecule FRET experiment.

5. **Dynamics Timscales:** The information about the timescales of the interdomain motions of PGK is still missing in the full time-interval. In fact, neutron spin echo spectroscopy (NSE) does not cover the time range above $\sim 100\,\mathrm{ns}$, and the analysis of the 2D plots gives only a rough estimate of the characteristic time of the interdomain motions. Furthermore, the experiments on diffusing molecules are limited to an upper time regime of a few milliseconds at normal experimental conditions. The characteristic

times of PGK interdomain dynamics can be obtained by single-molecule FRET-FCS on the microsecond to millisecond timescale [99], or by filtered FRET-FCS on the picoseconds to millisecond time interval [100][101]. The motions on longer timescales can only be assessed by increasing the observation time, which requires immobilized molecules and a different detection system, like, for instance, a total internal reflection microscope (TIRF).

6. **Improved Measurements:** The average number of photons detected during each single-molecule event can be improved by increasing the performances of the detection system, and /or by means of photoprotection systems [102][103][104] that must be optimized for the employed FRET pair (Alexa488/Alexa647). For example, a blue excitation at 485 nm would increase the number of collected photons by circa 20 %. Moreover, the methodology to perform time-resolved anisotropy measurements on labeled molecules is currently under study [105][106]. This will give more information about the environment surrounding the labeling sites, and about the dyes rotational freedom.

7. **Model Refinement:** The elastic network model must be refined by using a larger number of inter-dye distances. Furthermore, Multiparticle Collision Dynamics simulations [31] must be performed in order to get the timescale of the interdomain motions within the partially closed state. In addition, the velocity field of the solvent can be calculated, showing the interplay between the solvent flux induced by the protein motions and the substrate diffusion in proximity of the binding sites. Furthermore, the possible interaction between the dyes and the substrates can be analyzed and sorted out.

Appendices

Appendix A

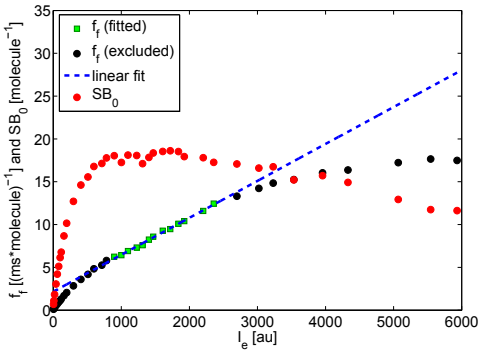
Single-Molecule Detection in Solution

A.1 Power Series Measurements

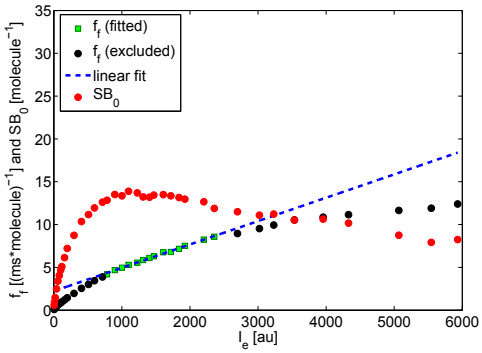
The power series, which were employed to determine the saturation intensity I_s, were measured as described in following. Whenever proteins were used, the cover slides were treated with Sigmacote®to prevent losses. Otherwise, slides cleaned with ethanol and acetone HPLC grade for 10 min each were utilized. The sample holder was sealed with Parafilm to prevent evaporation. The sample concentration was adjusted to N between 2 and 5 in order to have good signal-to-noise ratio in the FCS measurements (see Sec.5.3), and to avoid APDs saturation at high excitation intensities. The laser intensity was changed from 5 to 6000 arbitrary units, and for each step, a short FCS measurement (30 s) was carried out. Long FCS measurements (10 min) were performed before and after the power series, to determine N and to check weather the sample stuck to the slide. An average decrease of 22 % in the fluorophore concentration was observed for all samples, except for free dyes. The total count rates s_f $\left[\frac{\text{counts}}{\text{ms}}\right]$ were determined for each step, by division of the total number of detected photons N_{ph} and the total measuring time t_{tot} [ms]. The same procedure was followed for the buffer solution as a means to measure the background count rate bg_f $\left[\frac{\text{counts}}{\text{ms}}\right]$. The molecular brightness f_f $\left[\frac{\text{counts}}{\text{ms}\cdot\text{molecule}}\right]$ was calculated with Eq.4.8. The signal-to-background ratio SB_0 was obtained with Eq.4.10. Finally, the results were plotted against the laser intensity (Fig.A.1, Fig.A.2, Fig.A.3 and Fig.A.4). As stated in [42], the linear dependence between f_f and I_e corresponds to a plateau in the (SB_0, I_e) data set. The (f_f, I_e) data were then fitted with a straight line, starting from the beginning of the plateau until the fit worsened. The point of the fit at highest laser intensity corresponds to the saturation intensity (Tab.A.1).

Table A.1: Measured saturation intensities I_{sat} for single-labeled DNA, PGK, Strp and free Atto dyes. 10bp and 17bp refers to the base pair distance between the two dyes in the corresponding dlDNA FRET samples. A=Alexa647 and D=Alexa488.

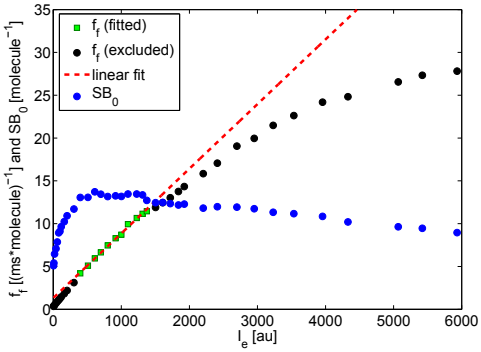
sample	I_s[a.u.]	sample	I_s[a.u.]	sample	I_s[a.u.]
A-slDNA 10bp	2400	A-slPGK Q135C	1900	A-Strp	2000
A-slDNA 17bp	2400	A-slPGK S290C	1900	D-Strp	700
D-slDNA 10bp	1400	D-slPGK Q135C	500	Atto655-NHS	2700
D-slDNA 17bp	1400	D-slPGK S290C	600	Atto488-NHS	2150



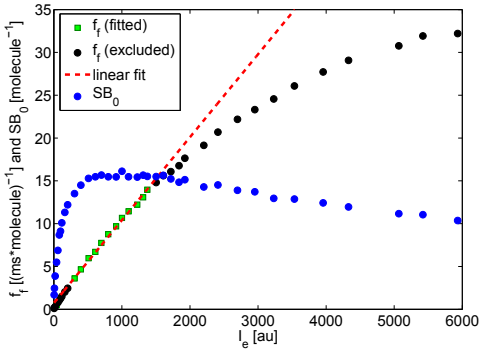
(a) A-slDNA 10bp



(b) A-slDNA 17bp

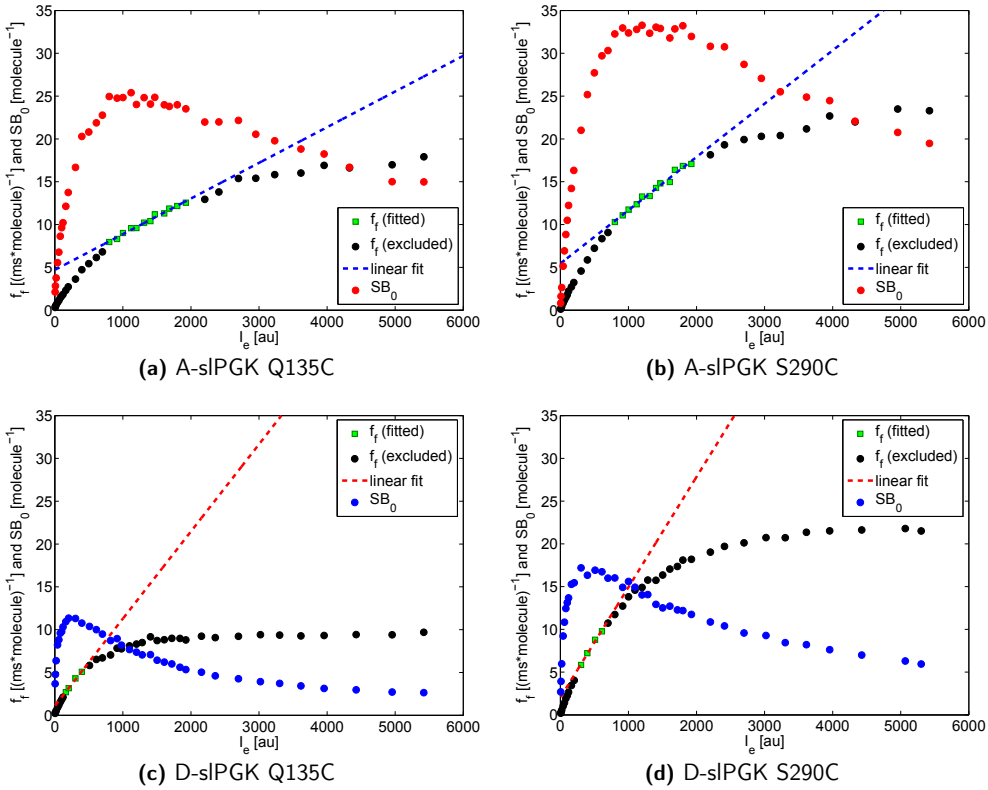


(c) D-slDNA 10bp



(d) D-slDNA 17bp

Figure A.1: Measured power series: slDNA (D=Alexa488 and A=Alexa647).



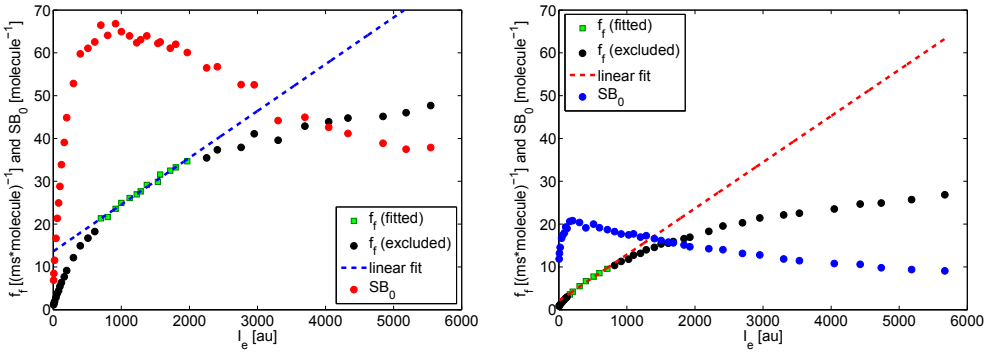
(a) A-slPGK Q135C

(b) A-slPGK S290C

(c) D-slPGK Q135C

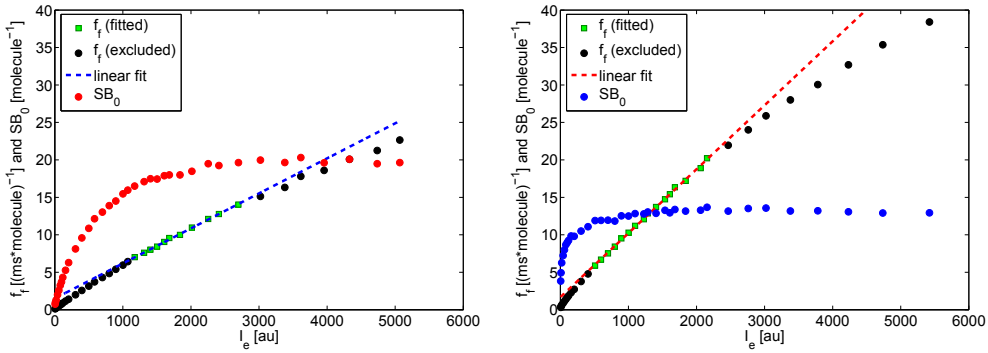
(d) D-slPGK S290C

Figure A.2: Measured power series: slPGK (D=Alexa488 and A=Alexa647).



(a) A-Strp **(b)** D-Strp

Figure A.3: Measured power series: slStrp (D=Alexa488 and A=Alexa647).



(a) Atto655-NHS **(b)** Atto488-NHS

Figure A.4: Measured power series: Atto488-NHS and Atto655-NHS

Appendix B

Fluorescence Lifetime

B.1 Single-Molecule Lifetime Estimator

The two equations reported in Sec.6.4 are derived here. By definition, the lifetime τ is the average time $\langle t \rangle$ spent by a fluorescent molecule in the excited state from the excitation until the photon emission, and, again by definition, the average time $\langle t \rangle$ is the mean of the distribution of times $p(t)$:

$$\tau \equiv \langle t \rangle \equiv \int t \cdot p(t)\, dt \tag{B.1}$$

Where $p(t)$ is the normalized fluorescence intensity decay $I(t)$:

$$p(t) = \frac{I(t)}{\int I(t)\, dt} \tag{B.2}$$

If the delay time δt is given by the time interval between the fluorescence excitation at time $t_0 = 0$ and the photon emission at time t ($\delta t = t - t_0 = t$), the delay times δt are distributed like the time t: $p(\delta t) = p(t)$. Therefore the average delay time $\langle \delta t \rangle$ is equivalent to the average time spent in the excited state $\langle t \rangle$ and to the lifetime τ ($\langle \delta t \rangle \equiv \langle t \rangle \equiv \tau$). At single molecule level, the distribution of time $p(t) = p(\delta t)$ cannot be retrieved from the intensity decay $I(t)$ with Eq.B.2 like at ensemble level. Therefore, the mean of the distribution must be estimated differently. That being said, since the detected photons are uncorrelated [74], the mean $\langle \delta t \rangle$ of the delay time distribution $p(\delta t)$ and the expected single molecule lifetime $\tau_{sm} = \langle \delta t \rangle$ are calculated as the sample mean of the measured delays δt_i (Eq.6.11 in the main text):

$$\tau_{sm} = \langle \delta t \rangle = \frac{1}{N_{ph}} \sum_{i=1}^{N_{ph}} \delta t_i \tag{B.3}$$

Where, N_{ph} is the total number of photons S_f detected in the course of a single-molecule

crossing event. The lifetime estimator given in Eq.B.3 is consistent [74], meaning that the lifetime estimate τ_{sm} converges towards the true value τ for a high number of detected photons N_{ph}.

To derive Eq.6.12 in the main text, the distribution p(t) must be explicitly written. Here p(t) is obtained by modeling the intensity decay with a multiexponential function $I(t) = F_{ME}(t) = \sum a_i \cdot e^{-\frac{t}{\tau_i}}$. Under this hypothesis, the distribution of times p(t) (Eq.B.2) becomes multiexponential:

$$p(t) = \frac{\sum a_i \cdot e^{-\frac{t}{\tau_i}}}{\int \sum a_i \cdot e^{-\frac{t}{\tau_i}} dt} \tag{B.4}$$

The average time $\langle t \rangle$ is then calculated using Eq.B.1 and Eq.B.4:

$$\langle t \rangle = \frac{\int t \cdot \left[\sum a_i \cdot e^{-\frac{t}{\tau_i}} \right] dt}{\int \sum a_i \cdot e^{-\frac{t}{\tau_i}} dt} \tag{B.5}$$

The sum $\sum a_i$ can be taken out from the integral, and the rule for the addition of integrals can be applied. The resulting integrals are equal to τ_i at the denominator, and following integration by parts to τ_i^2 at the numerator. The result is then by definition the intensity averaged lifetime $\langle \tau \rangle_f$ given by Eq.6.5:

$$\langle t \rangle = \frac{\sum a_i \cdot \tau_i^2}{\sum a_i \cdot \tau_i} = \langle \tau \rangle_f \tag{B.6}$$

Finally, since $\tau_{sm} = \langle \delta t \rangle = \langle t \rangle$, the equivalence (Eq.6.12) in the main text is attained:

$$\tau_{sm} = \langle \tau \rangle_f \tag{B.7}$$

B.2 Single-Molecule Lifetime Estimator with Multiple Detection Channels

In this work, the lifetime τ_{sm} was calculated with Eq.B.3 for each single molecule event. The measured delay times δt_i and the total number of photons N_{ph} recorded in each fluorescence burst were used. The delay times were calculated as $\delta t_i = t_i - t_0$ for each i-photon. Here, t_0 is the maximum in the TCSPC histogram built up with all the photons recorded by one detector in the course of the experiment. If two detectors (j = 1, 2) were used to record photons of the same color, the expected total lifetime $\bar{\tau}_{sm}$ was calculated with a weighted sum of the lifetimes $\tau_{sm,j}$, which were obtained for each j-channel by mean of Eq.B.3:

$$\bar{\tau}_{sm} = \frac{\sum w_j \cdot \tau_{sm,j}}{\sum w_j} \tag{B.8}$$

The weight w_j is the inverse of the standard deviation σ_j calculated for a single-molecule event in the j-channel:

$$w_j = \frac{1}{\sigma_j} = \left[\frac{1}{N_{ph,j} - 1} \cdot \sum_{i=1}^{N_{ph,j}} (\delta t_i - \langle \delta t \rangle_j)^2 \right]^{-\frac{1}{2}} \tag{B.9}$$

Here, $N_{ph,j}$ is the total number of photons detected in the j-channel during an individual burst. Eq.B.8 was used to calculate the lifetime τ from the estimated parallel and perpendicular lifetime components $\tau_{\perp,\parallel}$.

B.3 Single-Molecule Lifetime Bias Correction

The arithmetic mean is not a robust estimator [74]. Therefore, outliers can easily bias the calculated mean. Because of that, the robustness of the estimated single-molecule lifetime values τ_{sm} was tested by comparison of the best estimated values $\bar{\tau}_{sm}$ with the true values τ for a set of 6 dyes. The intensity averaged lifetime measured at ensemble level was used as the true value $\tau = \langle \tau \rangle_f$. The best estimate of the single-molecule lifetimes $\bar{\tau}_{sm}$ was retrieved with a Gaussian fit of the histogram which was obtained by grouping the lifetimes τ_{sm} from a set of bursts. Time windows of different widths were applied to the TCSPC histogram to filter out the photons from the single-molecule experiment (see Sec.7.2). The starting point of the time window was fixed at the rising edge of the IRF. The ending point was changed in order to have widths of 980, 1180 and 1380 time channels, respectively. The measured values of $\langle \tau \rangle_f$ vs. $\bar{\tau}_{sm}$ for the set of 6 dyes are displayed in Fig.B.1 and Tab.B.1 for the three employed time windows. Different behaviors are observed with respect to the expected linear relation $\bar{\tau}_{sm} = \langle \tau \rangle_f$ (Eq.B.7). At low lifetimes, the $\bar{\tau}_{sm}$ values are overestimated and the bias increases linearly with the width of the time window. Conversely at high lifetimes, the $\bar{\tau}_{sm}$ values are underestimated. A linear transformation was applied to correct the bias of the estimated lifetime τ_{sm}:

$$\tau'_{sm} = P \cdot \tau_{sm} + Q \tag{B.10}$$

Where, τ'_{sm} is the estimated unbiased single-molecule lifetime, and P and Q are the slope and the intercept of the linear transformation. Since, in the single-molecule FRET experiments conducted in this thesis, the interesting lifetime interval ends at ~ 3.3 ns (see Fig.10.2b), only the first 5 points were fitted to obtain P and Q. An optimal time window of 980 channels was choosen because the variance of the best value $\bar{\tau}_{sm}$ is lower with respect to longer intervals (data not shown). The best lifetimes, estimated with and without bias correction, are compared in Fig.B.2. The standard deviations are also shown. The relative deviation $\delta\tau = \bar{\tau}_{sm} - \langle \tau \rangle_f$ of the measured single-molecule best values $\bar{\tau}_{sm}$ from the true values $\langle \tau \rangle_f$ decreases after the linear

transformation. On the other hand, the standard deviation increases by $\sim 10\,\%$. A price must then be paid to get higher precision on the estimated single-molecule lifetimes. In the treated example, the photons were recorded with two SPADs for each color in order to collect quanta of light with different polarizations. As a consequence the recovered correction coefficients P and Q are optimal for single-molecule FRET measurements where the same set-up was used. Other configurations give different values.

Table B.1: Intensity averaged lifetime $\langle\tau\rangle_{\mathsf{f}}$ and best single-molecule lifetime estimate $\bar{\tau}_{\mathsf{sm}}$ shown as a function of the time interval (ΔChs $= 980/1180/1380$). The correction factors P and Q are reported for each interval. The single-molecule measurements were performed with a polarizer cube in order to measure the parallel and perpendicular component for each color. The asterisk (*) referes to quenched maleimide functionalized Alexa dyes wrongly stored in DMSO at $-20\,^{\circ}$C for a few weeks. The percentage relative deviation of the biased $\bar{\tau}_{\mathsf{sm}}$ and unbiased $\bar{\tau}'_{\mathsf{sm}}$ estimated single-molecule lifetime with respect to the expected value $\langle\tau\rangle_{\mathsf{f}}$ are also shown. The values are in [ns].

sample	$\langle\tau\rangle_{\mathsf{f}}$	$\bar{\tau}_{\mathsf{sm}}^{980}$	$\bar{\tau}_{\mathsf{sm}}^{1180}$	$\bar{\tau}_{\mathsf{sm}}^{1380}$	$\delta\tau_{980}\,(\%)$	$\delta\tau'_{980}\,(\%)$
Al647*-mal	0.83	0.95	1.04	1.17	$+14.5\,\%$	$-1.2\,\%$
Al647-mal	1.04	1.12	1.21	1.35	$+7.7\,\%$	$-3.9\,\%$
Atto655-NHS	1.79	1.89	2.01	2.11	$+5.6\,\%$	$+2.8\,\%$
DA-dlDNA 17bp	2.58	2.64	2.77	2.86	$+2.3\,\%$	$+7.4\,\%$
Al488*-mal	3.52	3.36	3.55	3.68	$-4.6\,\%$	$-2.0\,\%$
Al488-mal	3.83	3.49	3.69	3.82	$-8.9\,\%$	$-6.3\,\%$
P	/	1.095	1.053	1.061	/	/
Q	/	-0.2249	-0.2756	-0.4181	/	/

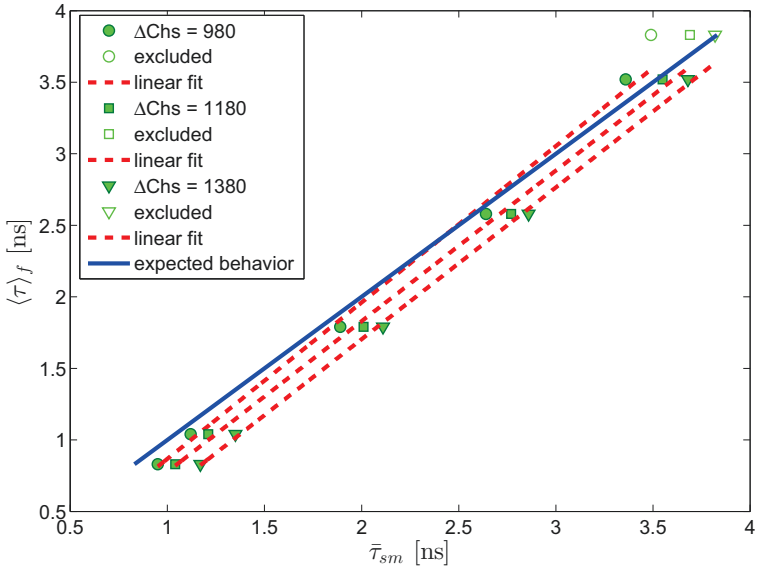
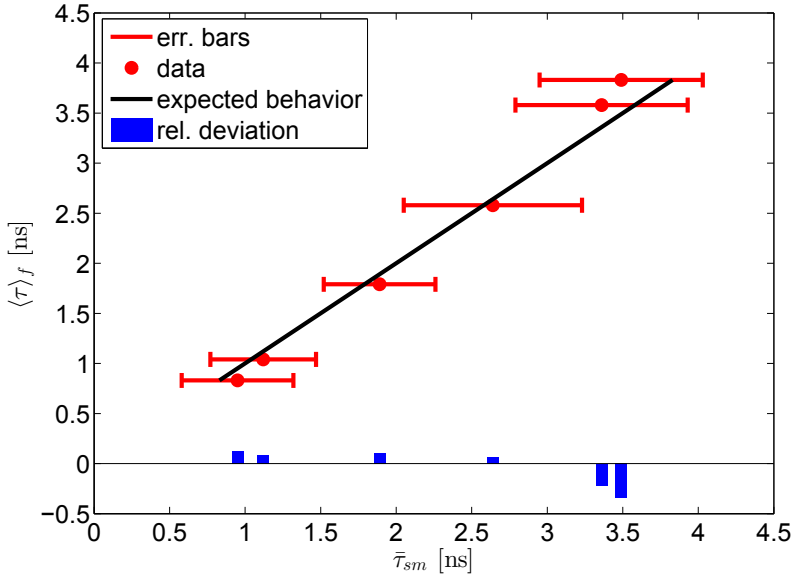
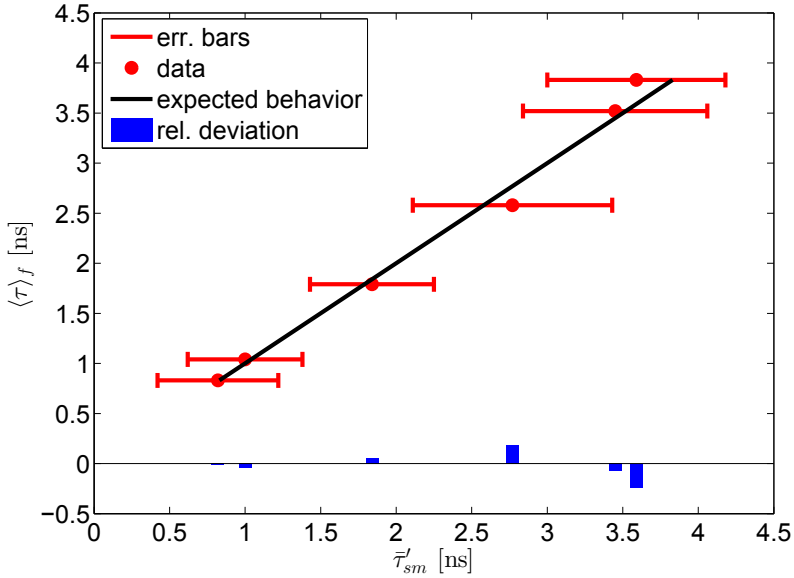


Figure B.1: Linear fits of the $\langle\tau\rangle_f$ vs. $\bar{\tau}_{sm}$ data. Time windows with different widths (ΔChs) were used. The points at higher lifetimes were excluded from the fit. The polarizer cube was used for the single-molecule measurements.



(a) P=0, Q=0



(b) P=1.095, Q=-0.2249

Figure B.2: Comparison of the best single-molecule lifetime estimated values before (B.2a) and after (B.2b) the bias correction. The standard deviations and the relative deviations $\delta\tau = \bar{\tau}_{\mathsf{sm}} - \langle\tau\rangle_{\mathsf{f}}$ with respect to the true values are shown. The polarizer cube was used for the single-molecule measurements.

Appendix C

Analysis of smFRET Measuraments with 1D and 2D Frequency Histrograms

C.1 Derivation of Eq.8.4

To derive Eq.8.4 in the main text, the fluorescence detection rates f_D and f_A must be written under FRET coupling. If a three state energy scheme (Fig.6.1) is considered, f_D and f_A are [53]:

$$f_D = p_{ex} \cdot g_D \cdot \phi_D \cdot \frac{k_{D0}}{k_{D0} + k_{ET}} \tag{C.1}$$

$$f_A = p_{ex} \cdot g_A \cdot \phi_A \cdot \frac{k_{ET}}{k_{D0} + k_{ET}} \tag{C.2}$$

In Eq.C.1, the first term p_{ex} is the probability to excite the donor that is equal to the product of the laser intensity I_e and the absorption cross section $\sigma_{01}(\lambda_{ex})$ (Eq.4.1). The other terms are: the probability that the donor decays to the ground state $\left(\frac{k_{D0}}{k_{D0}+k_{ET}}\right)$, it does so by emitting a photon (ϕ_D) (Sec.9.3.1), and the photon is detected (g_D). For the acceptor, a similar arguments holds (Eq.C.2), the only difference is the probability to receive energy from the excited donor through the FRET channel $\left(\frac{k_{ET}}{k_{D0}+k_{ET}}\right)$. To derive the desired equation (Eq.8.4), the equivalence $\left(\frac{k_{D0}}{k_{D0}+k_{ET}} = 1 - \frac{k_{ET}}{k_{D0}+k_{ET}}\right)$, and the definition $\left(E \equiv \frac{k_{ET}}{k_{D0}+k_{ET}}\right)$ (Eq.8.1) are first used:

$$f_D = p_{ex} \cdot g_D \cdot \phi_D \cdot (1 - E) \tag{C.3}$$

$$f_A = p_{ex} \cdot g_A \cdot \phi_A \cdot E \tag{C.4}$$

Then, the ratio $\frac{f_D}{f_A}$ between the rates is calculated, and the definition of fluorescence intensity ($F_f = f_f \cdot T$) is used (Eq.4.3). After simplification and some rearrangement Eq.8.4 is obtained:

$$E = \left[1 + \frac{f_D}{f_A}\right]^{-1} = \frac{F_A}{F_A + \gamma \cdot F_D} \tag{C.5}$$

The correction factor $\gamma = \gamma' \cdot g'$ is defined as the product of two ratios, the quantum yield ratio $\gamma' = \frac{\phi_A}{\phi_D}$ and the detection efficiency ratio $g' = \frac{g_A}{g_D}$.

C.2 Derivation of Eq.8.8

An explicit relation between the $\frac{f_D}{f_A}$ ratio and the physical parameters of the system is derived using Eq.C.1 and Eq.C.2 [53]:

$$\frac{f_D}{f_A} = \frac{\phi_D \cdot g_D \cdot k_{D0}}{\phi_A \cdot g_A \cdot k_{ET}} \tag{C.6}$$

The dependence from the physical parameters is obtained from the following definitions $k_{D0} = \frac{1}{\tau_{D0}}$, $k_{ET}(r) = \frac{1}{\tau_{D0}} \cdot \left(\frac{R_0}{r}\right)^6$, and using the definition of R_0 (Eq.8.3):

$$\frac{F_D}{F_A} = \frac{f_D}{f_A} = \frac{r^6 \cdot n^4}{\phi_A \cdot g' \cdot k^2 \cdot J(\lambda)} \tag{C.7}$$

Eq.C.7 shows that the measured F' and E (Eq.8.4 and Eq.8.5) are independent from the donor quantum yield ϕ_D, and highligths which parameters may contribute to the variance $\sigma^2_{E/F'}$ of the FRET indicators distributions p(E) and p(F') (Sec.8.4).

C.3 Variance of the FRET Efficiency Distribution

In this section, the relations employed to calculate the shot-noise σ^2_{SN} and the acceptor quantum-yield $\sigma^2_{\phi_A}$ contributions to the total variance $\sigma^2_E(\phi_A, r)$ of the efficiency distributions p(E) are reported. The dependence of the conformational variance σ^2_r from the underlying inter-dye distance distribution p[r(t)] is also presented.

C.3.1 Shot-Noise Variance: σ^2_{SN}

Shot-noise variance origins from the randomness of the photon emission and detection processes. For a diffusing molecule emitting an average number of photons $\langle N \rangle = \langle F_A + F_D \rangle$ higher than the threshold N_T ($\langle N \rangle >> N_T$), the shot-noise variance σ^2_{SN} is [63][64]:

$$\sigma^2_{SN} = \frac{\langle E \rangle \cdot (1 - \langle E \rangle)}{\langle N \rangle} \tag{C.8}$$

Where, $\langle E\rangle \equiv \int E \cdot p(E)\, dE$ is the mean of the FRET efficiency distribution p(E). The shot-noise variance σ^2_{SN} is limited by an upper value if the average number of photons $\langle N\rangle$ is lower than the threshold N_T ($\langle N\rangle << N_T$):

$$\sigma^2_{SN} < \frac{\langle E\rangle \cdot (1 - \langle E\rangle)}{N_T} \tag{C.9}$$

In this work, the mean efficiency $\langle E\rangle$ was obtained by fitting the FRET efficiency distributions recovered from the smFRET experiments with a multi gaussian model function (Eq.8.11). Moreover, the average number of photons $\langle N\rangle$ was calculated as follows:

$$\langle N\rangle \equiv \int N \cdot p(N)\, dN = \frac{\sum N \cdot \mathrm{hist}(N)}{\sum \mathrm{hist}(N)} \tag{C.10}$$

Here, p(N) is the theoretical distribution of the total number of photons N detected in individual fluorescence burst, and hist(N) is the experimental distribution obtained by histogramming the set of experimental values $\{N_i\} = \{(F_A + F_D)_i\}$.

C.3.2 Acceptor Quantum Yield Variance: $\sigma^2_{\phi_A}$

To derive the variance $\sigma^2_{\phi_A}$ of the acceptor quantum yield distribution $p(\phi_A)$ (Eq.8.10), the following equation is used as a starting point [68]:

$$\sigma^2_{\tilde{r}} = \langle r_{DA}\rangle_E \cdot \langle \phi_A\rangle^{\frac{1}{3}} \cdot \left[\langle (\phi_A^{-\frac{1}{6}})^2\rangle - \langle \phi_A^{-\frac{1}{6}}\rangle^2\right] \tag{C.11}$$

Where, $\sigma^2_{\tilde{r}}$ is the variance of the distribution $p(\tilde{r})$ of the apparent FRET averaged inter-dye distance $\tilde{r}$, and $\langle r_{DA}\rangle_E$ is the real FRET averaged inter-dye physical distance. Moreover, $\langle \phi_A\rangle$ is the mean acceptor quantum yield and $\left[\langle (\phi_A^{-\frac{1}{6}})^2\rangle - \langle \phi_A^{-\frac{1}{6}}\rangle^2\right]$ is the variance of $\phi_A^{-\frac{1}{6}}$. This last two terms are calculated as described in Appx.E.3.3 from the lifetime components $\{\tau_{A,i}\}$ of the acceptor fluorescent decay fitted with a multiexponential model function $F_{ME}(t)$ (Sec.6.3.1). The physical distance $\langle r_{DA}\rangle_E$ is also derived as shown in Appx.E.3.3. The variance $\sigma^2_{\tilde{r}}$ is then a measurable quantity. From this, the variance contribution $\sigma^2_{\phi_A}$ is calculated assuming that the function E(r) (Eq.8.2) is antisymmetric $[E(\langle\tilde{r}\rangle + \delta\tilde{r}) = -E(\langle\tilde{r}\rangle - \delta\tilde{r})]$ for small fluctuations around the mean apparent inter-dye distance $\tilde{r}$. The following expression is used:

$$\sigma^2_{\phi_A} \simeq \frac{1}{2} \cdot [E^+ - E^-]^2 \tag{C.12}$$

Where, the upper and lower values $E^{\pm} = \langle E\rangle \pm \Delta E$ were calculated by plugging the distances $r^{\mp} = \langle\tilde{r}\rangle \mp \sigma_{\tilde{r}}$ in Eq.8.2. In order to do so, the average apparent distance $\langle\tilde{r}\rangle$ was calculated with Eq.8.2 from the mean transfer efficiency $\langle E\rangle$ obtained from the fit of the efficiency

distributions p(E), and the standard deviation $\sigma_{\tilde{r}}$ was calculated with Eq.C.11.

C.3.3 Conformational Variance: σ_r^2

Conformational variance σ_r^2 can origin from inter-dye distance fluctuations or from a broad quasi-static distance distribution [63]. In the most general case, a time dependent distance distribution p[r(t)] is considered giving the following expression for the conformational variance σ_r^2:

$$\sigma_r^2(T) = \frac{2}{T^2}\int_0^T (T-t)\langle \delta E[r(t)] \cdot \delta E[r(0)]\rangle \, dt \tag{C.13}$$

Eq.C.13 shows that the conformational variance $\sigma_r^2(T)$ depends on the autocorrelation function of the efficiency fluctuations ($\delta E[r(t)] \equiv E[r(t)] - \langle E\rangle$) induced by the distance changes during the observation time T. The previous expression simplifies for relaxation times τ_c of the autocorrelation function much shorter ($\tau_c << T$) or much longer ($\tau_c >> T$) than the observation time. In the first case, for distance fluctuations much faster than the observation times ($\tau_c << T$), an approximated relation holds:

$$\sigma_r^2(T) \simeq \frac{2\tau_c \cdot \langle \delta E^2\rangle}{T} \tag{C.14}$$

In the second case, when the distance fluctuations are far slower than the observation time ($\tau_c >> T$), and, consequentely, a quasi-static distribution $p[r(t)] \sim p(r)$ can be considered, Eq.C.13 simplifies:

$$\sigma_r^2 \equiv \langle \delta E^2\rangle \equiv \int (E(r) - \langle E\rangle)^2 \cdot p(r)\, dr \tag{C.15}$$

If the dynamics are fast ($\tau_c << T$), the conformational variance change with the observation time T (Eq.C.14). If the dynamics are slow ($\tau_c >> T$), the variance σ_r^2 depends on the shape and extension of the quasi-static distance distribution p(r) (Eq.C.15). Therefore, static and dynamic distributions can be recognized by changing the time T in the framework of binwise analysis [69][70][72][71]. Furthermore, the conformational variance disappears if Eq.C.14 or Eq.C.15 are zero. The first equation (Eq.C.14) drops to zero ($\sigma_r^2(T) \to 0$) when the dynamics are very much faster than the observation time. The second equation (Eq.C.15) vanishes when the distribution of inter-dye distances resembles a delta function $p(r) \sim \delta(r - r_{DA})$. This last situation corresponds to a well defined conformation of the system. As a consequence, the two situations cannot be distinguished from the measured FRET efficiency distributions p(E) because they are degenerate ($\sigma_r^2 \simeq \sigma_r^2(T) \simeq 0$).

C.4 "Corrected" Static Line and Dynamic Line

The discrete set of amplitude $\{\langle\tau\rangle_{x,j}\}$ and intensity $\{\langle\tau\rangle_{f,j}\}$ averaged lifetimes, which must be fitted to derive the coefficients $\{c_i\}$ of the empirical polynomial transformation described in Sec.8.6.2, were calculated using the definitions of amplitude and intensity averaged lifetime (Eq.6.7 and Eq.6.5) generalized to a continuous distribution of distances p(r) [72]:

$$\langle\tau\rangle_x = \frac{\int p(r)\cdot\tau(r)\,dr}{\int p(r)\,dr} \tag{C.16}$$

$$\langle\tau\rangle_f = \frac{\int p(r)\cdot\tau(r)^2\,dr}{\int p(r)\cdot\tau(r)\,dr} \tag{C.17}$$

Where, $\tau(r) = \langle\tau_{D0}\rangle_f \cdot \left[1+\frac{R_0}{r}\right]^{-6}$ (Eq.6.2) defines the dependence of the lifetime from the inter-dye distance r. To perform the full procedure, the Förster radius R_0 (Sec.9.6), the distance distribution p(r), and the donor only intensity averaged lifetime $\langle\tau_{D0}\rangle_f$ (Sec.9.4) must be measured. What differentiates the calculation of the "corrected" static line ($\tau_c << T$) from the calculation of the dynamic line ($\tau_c \sim T$) are the functions used to model the distance distribution p(r).

C.4.1 "Corrected" Static Line

For fast fluctuations ($\tau_c << T$), the distribution of distances p(r) is well approximated by an empirical normalized gaussian function $p(r) \simeq p_G(r)$ that can be used to calculate the set of $\{\langle\tau\rangle_{x,j}\}$ and $\{\langle\tau\rangle_{f,j}\}$ values [33][72]:

$$p_G(r) = \frac{1}{\sqrt{2\pi}\cdot\sigma_{DA}}\cdot e^{-\frac{[r-\langle r_{DA}\rangle)]^2}{2\sigma_{DA}^2}} \tag{C.18}$$

The standard deviation σ_{DA} was measured from the fit of the donor eTCSPC histogram (Sec.6.3.2) and fixed during the whole procedure. Afterwards, a series of gaussian functions $\{p_{G,j}(r)\}$ was generated with Eq.C.18 by varying the average distance $\{\langle r_{DA,j}\rangle\}$ from $(0.1\cdot R_0)$ to $(100\cdot R_0)$. Finally, the amplitude and intensity averaged lifetime values ($\{\langle\tau\rangle_{x,j}\}$ and $\{\langle\tau\rangle_{f,j}\}$) were calculated for each gaussian $p_{G,j}(r)$ of the set $\{p_{G,j}(r)\}$ with Eq.C.16 and Eq.C.17. To do so, a numerical integration was performed in the interval from $[\langle r_{DA,j}\rangle - 4\cdot\sigma_{DA}]$ to $[\langle r_{DA,j}\rangle + 4\cdot\sigma_{DA}]$.

C.4.2 2-State Dynamic Line

A weighted sum $p(r) \simeq p_{2G}(r)$ of two gaussian function $p_{G,i}(r)$ (i = 1,2) (see Eq.C.18) was used to model the fast dynamics superimposed on the slow interconversion between 2-states having FRET averaged distances $\langle r_{DA}\rangle_{E1}$ and $\langle r_{DA}\rangle_{E2}$:

$$p_{2G}(r) = x_1 \cdot p_{G,1}(r) + (1 - x_1) \cdot p_{G,2}(r) \tag{C.19}$$

Here, the weight x_1 defines the relative populations of the two states, and the standard deviation of the gaussians $\sigma_{DA} = \sigma_{DA,1} = \sigma_{DA,2}$ encloses the information about the fast dynamics of the system. The standard deviation σ_{DA} was obtained by fitting the measured donor eTC-SPC histograms (see Sec.6.3.2). The average positions of the two gaussians ($\langle r_{DA}\rangle_{E1}$ and $\langle r_{DA}\rangle_{E2}$) were calculated with Eq.8.2 from the mean FRET efficiencies ($\langle E_1\rangle$ and $\langle E_2\rangle$) of the two interconverting states. These two efficiency values were previously determined by fitting the 1D FRET efficiency histograms, as described in Sec.8.5. Then, a series of distributions $\{p_{2G,j}(r)\}$ were determined by varying x_1 from 0 to 1. Finally, the sets of amplitude $\{\langle\tau\rangle_{x,j}\}$ and intensity averaged lifetimes $\{\langle\tau\rangle_{f,j}\}$ were calculated with Eq.C.16 and Eq.C.17 for each distribution $p_{2G,j}(r)$.

Appendix D

Measurements of ϕ, τ_{D0}, α and g'

D.1 Cross-Calibration of the Reference Dyes

The cross-calibration to test the quantum-yield reference samples was performed as follow. The absorption spectra $OD(\lambda)$ were measured on the absorption spectrometer with the 50 mm path length cuvettes. The emission spectra $f(\lambda)$ were measured on the spectrofluorimeter with the 1.5 mm path length cuvette. The fluorescence emission was excited at $\lambda_{ex,B} = 470$ nm (blue dyes) and $\lambda_{ex,R} = 607$ nm (red dyes) in order to collect the photons in a wide wavelength range without cutting the spectra and at the same time to have a good signal. The temperature was kept between $(21 - 22)\,^{\circ}C$. The concentration series was prepared in the low optical density regime $[OD(\lambda) \cdot \ell \leq 0.01]$ in order to avoid inner filter effects [52] assuring the validity of the working equation (Eq.9.11). In fact, under this condition, the re-absorption probability of the emitted photons is below $100 \times (1 - 10^{-0.01}) = 2.3\,\%$. In order to do so, the measured absorption spectra $OD(\lambda)$ were always kept below an $OD_{max} = 0.01 \cdot \frac{50\,mm}{1.5\,mm} = 0.33$. Furthermore, band passes of $(1 - 2)$ nm were used to assure the monochromaticity of the excitation beam, and comparable band passes ($BP_{flu} \simeq BP_{abs}$) were set on the spectrofluorimeter and on the absorption spectrometer in order to measure the absorption and emission spectra with almost identical excitation intensities ($I_{flu} \simeq I_{abs}$). In addition, the same band passes were used for the unknown and the reference in order to fullfill the assumption of the working equation (Eq.9.13) ($I_U(\lambda_{ex}) \simeq I_R(\lambda_{ex})$). Finally, the band pass of the emission monochromator in the spectrofluorimeter was optimized to 8 nm for blue dyes and to 12 nm for the red dyes in order to collect as much photons as possible avoiding distorsions of the spectra.

The data analysis was performed as follows. The absorption spectra $OD(\lambda)$ were smoothed and the emission spectra $f(\lambda)$ were corrected for the wavelength dependent detection efficiency of the spectrofluorimeter. This was done as described in Appx.D.3. Then, the corrected emission spectra $f'(\lambda)$ were integrated, and the integral values $\{F_i\}$ were plotted against the correspond-

ing optical densities $\{OD_i(\lambda_{ex})\}$. Finally, the data were fitted with a linear relation where the intercept with the y-axis was fixed to zero ($q = 0$). This condition assures that the experiment is conducted in the low optical density regime. The slopes m obtained from the fits were used to cross-calibrate the quantum yields values ϕ_R of the dyes by using the previously measured refractive indeces n. The amplitude averaged lifetimes $\langle\tau_R\rangle_x$ were also measured on the same day and on the same samples used for the cross-calibration. This was done to univocally connect the lifetime of the dye to the quantum yield ϕ_R. This lifetime can then be used to assess the stability of the reference dye immediately before the quantum yield measurements on the confocal microscope.

D.2 Justification of Eq.9.7

For a fluorophore emdedded in a complex environment, the radiative (k_{10}) and non radiative (k_{nr}) decay rates can be distributed over a wide range of values $p(k_{10})$ and $p(k_{nr})$. Therefore, also the quantum yield and the lifetime, which are by definition function of the decay rates, are distributed as $p(\tau)$ and $p(\phi)$, and the measured values corresponds to the means $\langle\phi\rangle$ and $\langle\tau\rangle_x$. Told that, Eq.9.7 is derived from the definition of the average quantum yield $\langle\phi\rangle$ by considering a discrete distribution of states with relative populations x_i:

$$\langle\phi\rangle \equiv \int \phi \cdot p(\phi)\, d\phi \equiv \sum_i \phi_i \cdot x_i \tag{D.1}$$

The average quantum yield $\langle\phi\rangle$ is the sum of the quantum yield ϕ_i of individual states weighted by their relative population x_i. The previous expression can be rewritten by means of Eq.9.5 ($\tau_i(\phi) \equiv \tau_n \cdot \phi_i$):

$$\langle\phi\rangle \equiv \sum_i \frac{\tau_i(\phi) \cdot x_i}{\tau_n} \tag{D.2}$$

The final expression $\xi(\phi)$ is derived by recognizing that the definition of amplitude averaged lifetime ($\sum_i \tau_i(\phi) \cdot x_i = \langle\tau(\phi)\rangle_x$) appears in Eq.D.2:

$$\langle\tau(\phi)\rangle_x \equiv \tau_n \cdot \langle\phi\rangle \tag{D.3}$$

The working equation (Eq.9.7) is then easily obtained from the ratio $\frac{\langle\tau_R(\phi)\rangle_x}{\langle\tau_U(\phi)\rangle_x}$ assuming to have two dyes with similar natural lifetimes ($\tau_{n,R} \simeq \tau_{n,U}$):

$$\langle\phi_U^\tau\rangle \simeq \frac{\langle\tau_U(\phi)\rangle_x}{\langle\tau_R(\phi)\rangle_x} \cdot \langle\phi_R\rangle \tag{D.4}$$

D.3 Correction of the Emission Spectra

Corrected fluorescence emission spectra $f'(\lambda)$ must be known in order to calculate the quantum yields ϕ^f with the "brightness method" (Sec.D.4.2), in order to cross calibrate the reference quantum yields ϕ_R (Sec.9.3.2), and in order to obtain the Förster radius R_0 (Sec.9.6). Therefore, an empirical tranformation $O(\lambda)$ (Fig.D.1) was applied to the measured emission spectra $f(\lambda)$ to correct for the wavelength dependent detection efficiency of the spectrofluorimeter. The transformation was calculated as described in [52][107] with the following equation:

$$O(\lambda) = \frac{f^{\dagger}(\lambda)}{f(\lambda)} \tag{D.5}$$

Where, $f(\lambda)$ and $f^{\dagger}(\lambda)$ are the measured and the expected normalized emission spectra, respectively. Atto655-NHS and the reference dye sampler kit from Molecular Probes®were used to derive the transformation in a broad wavelength range $(365 - 845)$ nm. The emission spectra $f(\lambda)$ were measured at 22 °C for each reference dye with an excitation band pass of 1 nm. The expected emission spectra $f^{\dagger}(\lambda)$ were downloaded from the Invitrogen®and the Atto-Tech®web sites. The corrected spectra $f'(\lambda)$ were calculated as follow by considering

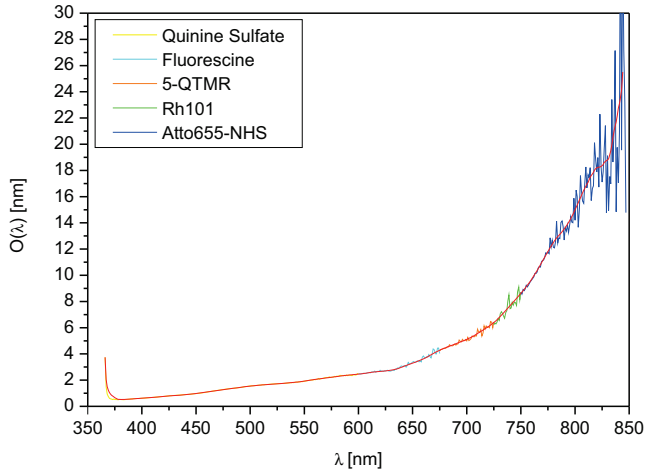
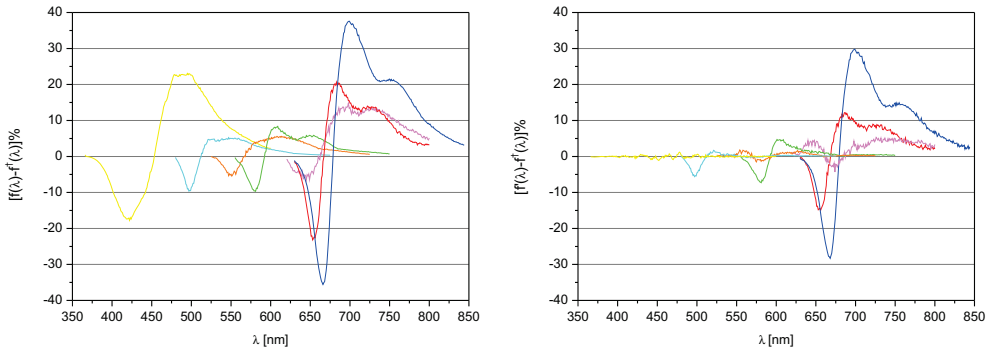


Figure D.1: Empirical transformation $O(\lambda)$ (red smoothed line) obtained with Eq.D.5 using the following dyes: Quinine Sulfate, Fluorescine, 5-QTMR, Rh101, Atto655-NHS.

that the transformation $O(\lambda)$ is not normalized:

$$f'(\lambda) = \frac{f(\lambda) \cdot O(\lambda)}{f(\lambda_{max}) \cdot O(\lambda_{max})} \cdot f(\lambda_{max}) \tag{D.6}$$

The relative deviations of the corrected $f'(\lambda)$ and uncorrected $f(\lambda)$ spectra from the expected one $f^{\dagger}(\lambda)$ were calculated for the reference dyes in order to verify the performances of the transformation $OD(\lambda)$ and to estimate the uncertainity of the measurement. As shown by Fig.D.2



(a) Before the correction: $[f(\lambda) - f^{\dagger}(\lambda)]\,\%$ **(b)** After the correction: $[f'(\lambda) - f^{\dagger}(\lambda)]\,\%$

Figure D.2: Relative deviation of the corrected $f'(\lambda)$ and uncorrected $f(\lambda)$ emission spectra from the expected spectra $f^{\dagger}(\lambda)$ for the following dyes: Atto655-NHS (blue), Al647-mal (red), Nile Blue (pink), Rh101 (green), 5-QTMR (orange), Fluorescine (cyan), Quinine Sulfate (yellow).

the transformation performs well. In fact, the relative deviations show a relevant reduction after the correction. Furthermore, a relative error below $5\,\%$ was observed in the interval from 345 nm to 600 nm. The relative error increases to circa $10\,\%$ at higher wavelengths where it also depends on the dye. For Al647 and Nile Blue a deviation below $10\,\%$ was observed while for Atto655-NHS the relative error increases to $(15 - 20)\,\%$. For this reason Atto655-NHS was not used for the cross calibration of the reference quantum yields ϕ_R (Sec.9.3.2).

D.4 Measurement of the Quantum Yield with the Confocal Microscope

In this work, the quantum yields of Al488 and Al647 bound to DNA and PGK were measured on the confocal microscope with two comparative methods, the "lifetime method" (Sec.D.4.1) and the "brightness method" (Sec.D.4.2). The measurements were performed on the single-labeled species (slPGK and slDNA). The previously cross-calibrated Al488-COOH and Al647-COOH free dyes were used as the reference samples (Sec.9.3.2), which means that $\phi_R^D = 0.92$ and $\phi_R^A = 0.33$ were taken as the reference quantum yields (Tab.9.1). Then, the unknown quantum yields of the double-labeled samples (dlPGK and dlDNA) were estimated with the

general strategies described in Sec.9.2. In addition, a correction to the quantum yields of double-labeled PGK was applied in order to consider the effects of the substrates (Mg_2ATP, KADP, 3-PG).

D.4.1 "Lifetime Method"

The "lifetime method" estimates the fluorescence quantum yield of the unknown ϕ_U^τ by means of Eq.9.7. In order to do so, the amplitude averaged lifetimes of the reference $\langle\tau_R\rangle_x$ and the unknown $\langle\tau_U\rangle_x$ must be measured. The measurement of the fluorescence lifetimes was performed as described in Sec.6.3, and the amplitude averaged lifetimes $\langle\tau\rangle_x$ were obtained with a multi-exponential model function $F_{ME}(t)$ (Sec.6.3.1). The reference free dyes and the unknown single-labeled samples (slPGK and slDNA) were measured on the same day and with the same optimal excitation intensities ($I_U = I_R$) (see Tab.4.2). The unknown quantum yields ϕ_U^τ calculated with Eq.9.7 are reported in Tab.D.1.

Table D.1: Measured amplitude averaged lifetimes $\langle\tau_U\rangle_x$ of Al488 and Al647 bound to single-labeled DNA ("10bp" and "17bp") and PGK (Q135C and S290C), and quantum yields ϕ_U^τ obtained with the "lifetime method". The notation used for the DNA samples, "10bp" and "17bp", refers to the distance between the donor and the acceptor binding positions. The following lifetimes and quantum yields were used as reference values: $\langle\tau_R^D\rangle_x = 4.04$ ns, $\phi_R^D = 0.92$, $\langle\tau_R^A\rangle_x = 0.83$ ns, $\phi_R^A = 0.33$. Single-labeled DNA was measured in the annealing buffer, slPGK in the FRET buffer and the reference dyes in PBS. A volume fraction of 0.001 % of Tween®20 was added to the buffer solution of PGK and DNA.

sample	$\langle\tau_U\rangle_x$ [ns]	ϕ_U^τ	sample	$\langle\tau_U\rangle_x$ [ns]	ϕ_U^τ
D-slDNA 10bp	3.71	0.85	D-slPGK Q135C	3.67	0.84
D-slDNA 17bp	3.69	0.85	D-slPGK S290C	3.54	0.81
A-slDNA 10bp	0.99	0.39	A-slPGK Q135C	1.19	0.47
A-slDNA 17bp	0.97	0.38	A-slPGK S290C	1.49	0.60

D.4.2 "Brightness Method"

The "brightness method" estimates the fluorescence quantum yield ϕ_U^f of the unknown sample by using Eq.9.9. Therefore, the molecular brightnesses f_R and f_U, the optical density spectra $OD_R(\lambda)$ and $OD_U(\lambda)$, and the optical detection efficiencies $g_{R,opt}$ and $g_{U,opt}$ must be measured for the reference and the unknown in order to perform the quantum yield measurements.
FCS measurements of 10 min were conducted on the single-labeled samples (slPGK and slDNA) at optimal conditions (see Sec.5.5) and with the excitation intensities reported in Tab.4.2, in order to obtain the total count rates s_f and the apparent average number of molecules N_{FCS}

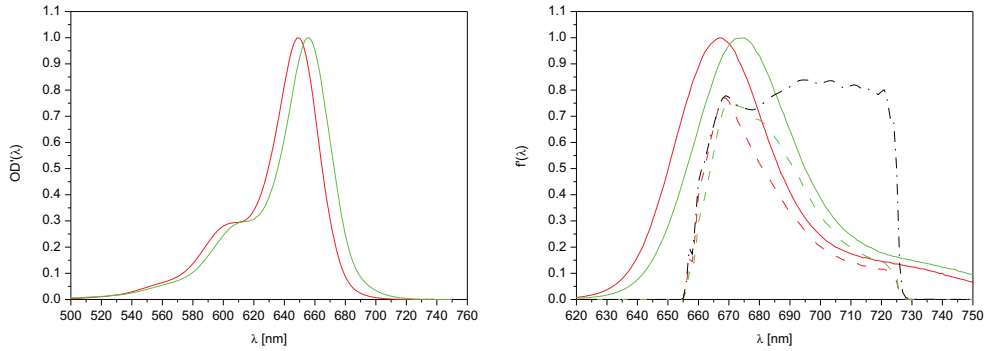
within the confocal detection volume. The background count rates bg_f were then measured with short FCS measurements of 2 min of the buffer only. Finally, the average number of molecules N corrected for the background contribution were calculated with Eq.5.7, and the molecular brightnesses f_f were estimated with Eq.4.8. The full procedure was repeated $4-5$ times in order to increase the precision of the measurement, and the arithmetic mean $\bar{f}_f$ of the set of measurements was taken as the molecular brightness best estimate.
The optical density spectra $OD(\lambda)$ of labeled PGK were measured on the absorption spectrometer during the labeling procedure, smoothed, and finally normalized at the maximum peak position $OD'(\lambda) = \frac{OD(\lambda)}{OD(\lambda_{max})}$ (see Fig.D.3a). On the other hand, the optical spectra $OD(\lambda)$ of labeled DNA cannot be measured because of the low amount of sample. Nonetheless, after some considerations, the ratio $\frac{OD'_R(\lambda_{ex})}{OD'_U(\lambda_{ex})}$ can be set to one (see Appx.D.5). The fluorescence emission spectra $f(\lambda)$ were measured on the spectrofluorimeter, corrected (see Appx.D.3), and then used to calculate the optical detection efficiencies g_{opt} by means of the total optical transmission function $T(\lambda)$ shown in Fig.4.1. The unknown quantum yields ϕ^f_U were then calculated with the working equation (Eq.9.9). The results are reported in Tab.D.2. An example of the measured $OD(\lambda)$ and $f'(\lambda)$ spectra is shown in Fig.D.3 to highlight the possible spectral shift induced by binding of the dye to the protein.

Table D.2: Measured quantum yields ϕ^f_U of slPGK and slDNA obtained with the "brightness method". The measured ratio between the optical detection efficiencies g_{opt} and the ratio of the normalized optical densities at the excitation wavelength $OD'(\lambda_{ex})$ are also shown. Single-labeled DNA was measured in the annealing buffer, slPGK in the FRET buffer, and the reference dyes in PBS. For slDNA and slPGK, 0.001 % of Tween®20 was added to the buffer solution. The following reference quantum yields were used: $\phi^D_R = 0.92$, $\phi^A_R = 0.33$. The notation used for the DNA samples, "10bp" and "17bp", refers to the distance between the donor and the acceptor binding sites. The full set of spectra is reported in Appx.D.5.

sample	$\frac{g_{R,opt}}{g_{U,opt}}$	$\frac{OD'_R(\lambda_{ex})}{OD'_U(\lambda_{ex})}$	ϕ^f_U	sample	$\frac{g_{R,opt}}{g_{U,opt}}$	$\frac{OD'_R(\lambda_{ex})}{OD'_U(\lambda_{ex})}$	ϕ^f_U
D-slDNA 10bp	1.013	~ 1	0.89	D-slPGK Q135C	0.994	1.183	0.88
D-slDNA 17bp	1.000	~ 1	0.88	D-slPGK S290C	1.005	1.159	0.77
A-slDNA 10bp	0.975	~ 1	0.37	A-slPGK Q135C	0.922	1.106	0.49
A-slDNA 17bp	0.978	~ 1	0.39	A-slPGK S290C	0.836	1.332	0.67

D.4.3 Best Estimate of the Quantum Yields of Single-Labeled Samples

The quantum yields of single-labeled samples (slDNA and slPGK), which were measured as described in the previous sections with the "lifetime method" ϕ^τ_U (Sec.D.4.1) and the "brightness method" ϕ^f_U (Sec.D.4.2), were then used to calculate the best estimate ϕ^{sl}_U of the donor



(a) **(b)**

Figure D.3: Measured spectra of A-slPGK (green) and free Al647-COOH (red). (a) Smoothed and normalized optical density spectra OD′(λ). (b) Corrected and normalized fluorescence emission spectra (continuos line), transmitted fluorescence spectra (dashed line) and total optical transmission function (black dashed line). All the measured spectra are reported in Appx.D.5.

and the acceptor quantum yields (ϕ_D^{sl} and ϕ_A^{sl}). Since the two methods are independent, and static quenching is not observed, the arithmetic mean was taken as the best estimate of ϕ_U^{sl}:

$$\phi_U^{sl} = \frac{\phi_U^{\tau} + \phi_U^{f}}{2} \tag{D.7}$$

In this way, the precision of the measurements increases. The calculated best estimates ϕ_U^{sl}, which are reported in Tab.D.3, show that the acceptor quantum yield ϕ_A^{sl} increases with respect to the unbound species ($\phi_R^A = 0.33$) when the dye binds to DNA or PGK, and that the quantum yields of the bound acceptor is higher on PGK with respect to DNA. This in turn, can be explained by the sterical shielding of the dye from the quenchers when Al647 binds to PGK or DNA, and by the higher molecular weight of PGK with respect to DNA [52]. In fact, it is known that the quenching rate k_q of bound fluorophores drops as a function of the biomolecules molecular weights. Nonetheless, a perturbation of the radiative decay rate k_{10}, which induces a change of the energy gap between the ground state S_0 and the lowest singlet state S_1 (Fig.6.1), can also explain the observed phenomenon (see Eq.9.4). Furthermore, the quantum yield ϕ_A^{sl} of the acceptor bound to PGK is strongly dependent on the binding position ($\phi_A^{Q135C} = 0.48$ and $\phi_A^{S290C} = 0.64$) while for DNA it is almost unchanged ($\phi_A^{10bp} \simeq \phi_A^{17bp} \simeq 0.39$). This indicates that PGK provides different environments for the acceptor on the two labeling positions, while on DNA, the labeling sites are equivalent. On the other hand, when Al488 binds to PGK or DNA, the donor quantum yields ϕ_D^{sl} slightly decrease with respect to the free dye ($\phi_R^D = 0.92$), and the quantum yield is almost independent from the binding position. A possible explanation

is the local quenching of the dye [108].

Table D.3: The quantum yields of the single-labeled samples (slDNA and slPGK) obtained with the "lifetime methods" ϕ_U^{τ} and the "brightness method" ϕ_U^{f} are shown together with the calculated best estimates ϕ_U^{sl} (Eq.D.7). The notation used for the DNA samples, "10bp" and "17bp", referes to the distance between the donor and the acceptor labeling sites.

sample	ϕ_U^{τ}	ϕ_U^{f}	ϕ_U^{sl}
A-slDNA 10bp	0.39	0.37	0.38
A-slDNA 17bp	0.38	0.39	0.39
D-slDNA 10bp	0.85	0.89	0.87
D-slDNA 17bp	0.84	0.88	0.86
A-slPGK Q135C	0.47	0.49	0.48
A-slPGK S290C	0.60	0.67	0.64
D-slPGK Q135C	0.83	0.82	0.83
D-slPGK S290C	0.81	0.77	0.79

D.4.4 The Quantum Yield Ratio: γ'

The following equations, which were derived by setting $\xi = \phi$ in the general expressions reported in Sec.9.2, were used to calculate the quantum yields ϕ_U of dlDNA and dlPGK by means of the quantum yields ϕ_U^{sl} of the single-labeled species:

$$\langle \phi_U \rangle = (0.36 \cdot \phi_U^{Q135C}) + (0.64 \cdot \phi_U^{S290C}) \tag{D.8}$$

$$\bar{\phi}_U = \frac{\phi_U^{10bp} + \phi_U^{17bp}}{2} \tag{D.9}$$

The obtained values are reported in Tab.D.4 with the resulting quantum yield ratios γ'.
In order to test the precision of Eq.D.8, the estimated quantum yield $\langle \phi_A \rangle$ of the acceptor bound to dlPGK, which is reported in Tab.D.4, was compared with the acceptor quantum yield directly measured on the double-labeled PGK sample with the lifetime method $\phi_A^{\tau,dl}$. In fact, the lifetime of the acceptor on the double-labeled samples can be measured upon direct excitation of the acceptor in order to avoid the energy transfer. The resulting amplitude averaged lifetime is $\langle \tau_A^{dl} \rangle_x = 1.33\,\text{ns}$ and corresponds to a quantum yield of $\phi_A^{dl} = 0.56$. The small relative deviation of $3.6\,\%$ between the estimated value of $\langle \phi_A \rangle = \langle 0.58 \rangle$ and the measured value $\phi_A^{\tau,dl} = 0.56$ indicates that the precision of the measurements is high. Hereby, a precise estimate of the donor quantum yield on the double-labeled sample $\langle \phi_D \rangle$ is also expected. In the main text the averaging notation $\langle .. \rangle$ will be discarded.

Table D.4: Estimated quantum yields ϕ_U and quantum yield ratios γ' of double-labeled DNA and PGK.

sample	ϕ_U	$\gamma' = \frac{\phi_A}{\phi_D}$
A-dlDNA	0.39	0.44
D-dlDNA	0.87	/
A-dlPGK	$\langle 0.58 \rangle$	0.73
D-dlPGK	$\langle 0.81 \rangle$	/

D.4.5 Correction for PGK's Substrate Effects

The quantum yields of the donor and the acceptor bound to double-labeled PGK, which are reported in Tab.D.4, must be corrected for the effects of PGK substrates (Mg_2ATP, KADP, 3-PG). In fact, the substrates alter the lifetimes and the quantum yields of the dyes. The "lifetime method" (Appx.D.4.1) was employed for the correction. Here, the quantum yields ϕ_U^S of dlPGK at different substrate conditions were corrected with the following equation with respect to the quantum yields ϕ_R^{NS} measured in pure FRET buffer (Tab.D.4: $\phi_A^{NS} = 0.58$ and $\phi_D^{NS} = 0.81$):

$$\phi_U^S \simeq \frac{\langle \tau_S \rangle_x}{\langle \tau_{NS} \rangle_x} \cdot \phi_R^{NS} \tag{D.10}$$

Where, $\langle \tau_{NS} \rangle_x$ and $\langle \tau_S \rangle_x$ are the amplitude averaged lifetimes of the double-labeled samples measured with (NS) and without (NS) the substrates. These two lifetimes were derived in a different way depending on the dye. For the donor, the lifetimes $\langle \tau_S^{sl} \rangle_x$ of the donor only species (Q135C and S290C) were first measured with the different substrates. Then, Eq.D.8 was used to estimate the lifetime $\langle \tau_S \rangle_x$ of the double-labeled species (Tab.D.5). For the acceptor, the lifetimes $\langle \tau_A^{dl} \rangle_x$ of the double-labeled sample were directly measured at different substrate conditions (Tab.D.6). As shown in red on the tables (Tab.D.5 and Tab.D.6), the fluctuations of the quantum yields, which are induced by the substrates around the arithmetic mean $\bar{\phi_U}$, are lower than the experimental errors ($\pm 0.02 \leq \pm 0.04$). Therefore, the donor and acceptor quantum yields ($\bar{\phi}_D = 0.81$ and $\bar{\phi}_A = 0.54$) averaged over the different experimental conditions were used to calculate the quantum yield ratio ($\gamma' = 0.67$) and to analyze the smFRET data. These last values are reported in the main text (Tab.9.2).

Table D.5: The amplitude average lifetimes $\langle\tau_S^{sl}\rangle_x$ of the donor only species (Q135C and S290C) measured at different substrate conditions (Mg_2ATP = KADP = 13 mM and 3-PG = 42 mM) and used to calculate the lifetimes of the double-labeled species are displayed. The correction factors and the corrected donor quantum yields ϕ_D^S are also shown. The donor quantum yield $\bar{\phi}_D$ averaged over the different experimental conditions is reported in red with its standard deviation. The lifetimes and the quantum yields measured without substrates (none) were used as the reference values to apply Eq.D.10: $\langle\tau_{NS}\rangle_x = 3.59$ ns (Tab.D.1), $\phi_D^{dl} = 0.81$ (Tab.D.4).

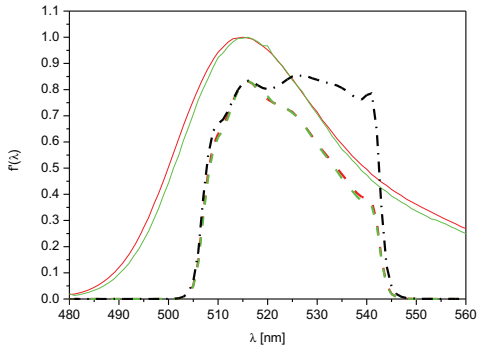
substrate	$\langle\tau_S^{sl}\rangle_x$ [ns]	$\langle\tau_S^{sl}\rangle_x$ [ns]	$\langle\tau_S\rangle_x$ [ns]	$\frac{\langle\tau_S\rangle_x}{\langle\tau_{NS}\rangle_x}$	ϕ_D^S
none	3.67	3.54	3.59	/	/
Mg_2ATP	3.53	3.44	3.47	0.994	0.79
KADP	3.52	3.51	3.51	1.006	0.80
3-PG	3.86	3.70	3.76	1.077	0.85
3-PG*KAPD	3.68	3.48	3.55	1.017	0.80
					$\langle 0.81 \pm 0.02\rangle$
sample	Q135C	S290C	dlPGK	dlPGK	dlPGK

Table D.6: The amplitude averaged lifetimes $\langle\tau_A^{dl}\rangle_x$ measured upon direct excitation of the acceptor bound to dlPGK at different substrates conditions (Mg_2ATP = KADP = 13 mM and 3-PG = 42 mM) are shown. The correction factor and the corrected acceptor quantum yields ϕ_A^S at different conditions are also reported. The acceptor quantum yield $\bar{\phi}_A$ averaged over the different experimental conditions is displayed in red with its standard deviation. The following lifetime and and quantum yield were used as reference values: $\langle\tau_{NS}\rangle_x = (0.36 \cdot 1.19\,\text{ns}) + (0.64 \cdot 1.49\,\text{ns}) = 1.38$ ns (Tab.D.1), $\langle\phi_D\rangle = 0.58$ (Tab.D.4).

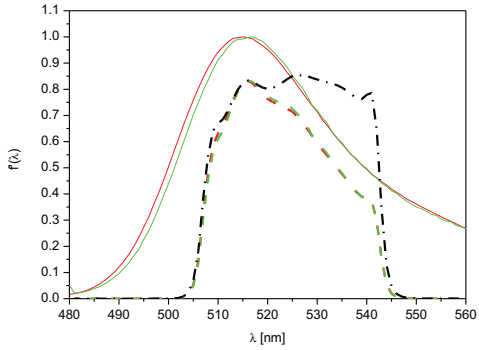
substrate	$\langle\tau_S\rangle_x$ [ns]	$\frac{\langle\tau_S\rangle_x}{\langle\tau_{NS}\rangle_x}$	ϕ_A^S
none	1.33	0.964	0.56
Mg_2ATP	1.23	0.891	0.52
KADP	1.28	0.928	0.54
3-PG	1.33	0.964	0.56
3-PG*KAPD	1.25	0.906	0.53
			$\langle 0.54 \pm 0.02\rangle$
sample	dlPGK	dlPGK	dlPGK

D.5 Measured Absorption and Emission Spectra for Quantum Yield Calculation

In this section, the normalized optical density spectra $OD'(\lambda)$, and the normalized corrected emission spectra $f'(\lambda)$ used to calculate the quantum yield of the dyes bound to slPGK and slDNA with the "brightness method" (see Sec.D.4.2) are reported. The transmitted fluorescence is also shown together with the total optical transmission function $T(\lambda)$ (see Sec.4.1.1). The absorption spectra of slDNA were not measured because of the low amount of available sample. Nevertheless, it is possible to assume that the optical density spectra of the acceptor bound to DNA are very similar to the spectra of the unbound dyes. In fact, from the spectra of slPGK (see Fig.D.6 and Fig.D.7) it is evident how the shifts of the emission spectra $f'(\lambda)$ and the absorption spectra $OD'(\lambda)$ with respect to the same spectra measured for the unbound dye are correlated. Therefore, even though the absorption spectra of slDNA cannot be measured, a relevant shift is really unfeasible because the measured emission spectra of the dye bound to DNA almost coincide to the one measured for the reference unbound dye (see Fig.D.4 and Fig.D.5). As a consequence, the $\frac{OD'_R(\lambda_{ecx})}{OD'_U(\lambda_{ex})}$ ratio can be safely approximated to 1 for the slDNA samples. The fluorescence was excited at 470 nm and 607 nm with a bandpass of 1 nm and 2 nm, and was measured with a band pass of 8 nm and 12 nm, respectively. The absorption was measured with a band pass of 1 nm. The DNA samples were measured in the Annealing buffer, the PGK samples in the FRET buffer, and the unbound dyes in PBS. A 0.001 $\%$ volume concentration of Tween®20 was added to the buffers used for the measurements of DNA and PGK.

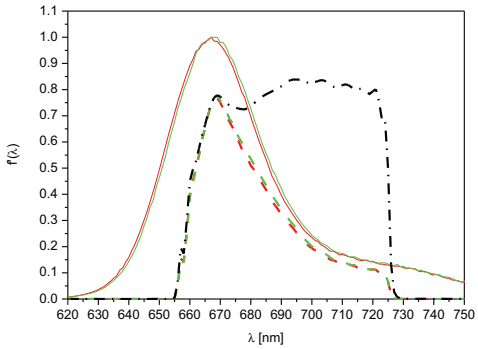


(a) D-slDNA 10bp

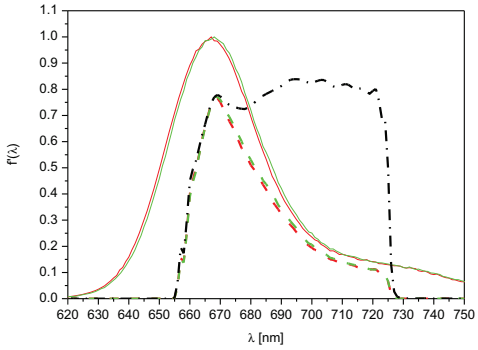


(b) D-slDNA 17bp

Figure D.4: Corrected and normalized (continuos line), and transmitted (dashed line) fluorescence emission spectra of D-slDNA (green) and unbound Al488-COOH (red). The total optical transmission function is also shown (black dashed line).

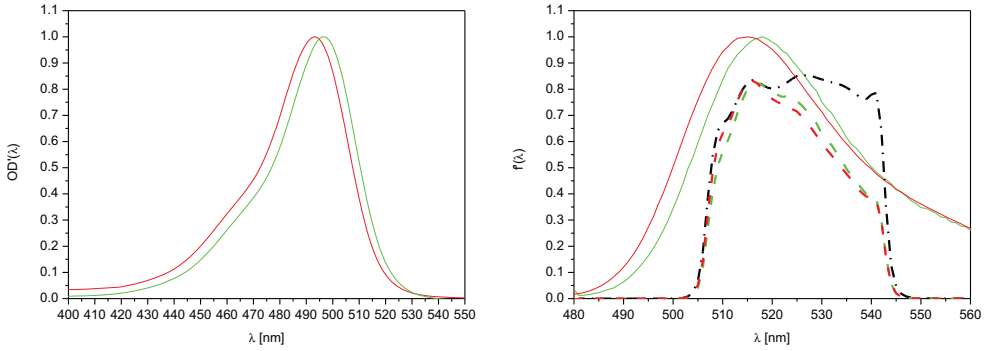


(a) A-slDNA 10bp



(b) A-slDNA 17bp

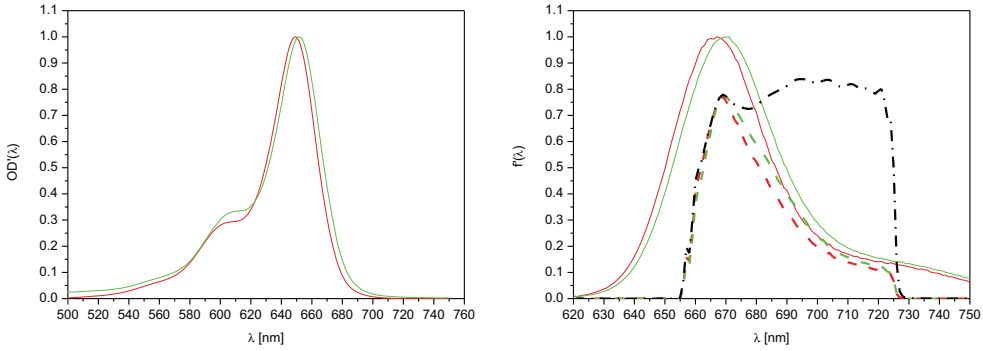
Figure D.5: Corrected and normalized (continuos line), and transmitted (dashed line) fluorescence emission spectra of A-slDNA (green) and unbound Al647-COOH (red). The total optical transmission function is also shown (black dashed line).



(a) D-slPGK Q135C **(b)** D-slPGK Q135C

(c) D-slPGK S290C **(d)** D-slPGK S290C

Figure D.6: 1) Left Column: smoothed and normalized optical density spectra OD′(λ) of D-slPGK (green) and unbound Al488-COOH (red). 2) Right Column: corrected and normalized (continuos line), and transmitted (dashed line) fluorescence emission spectra of D-slPGK (green) and unbound Al488-COOH (red). The total optical transmission function is also shown (black dashed line).



(a) A-slPGK Q135C **(b)** A-slPGK Q135C

(c) A-slPGK S290C **(d)** A-slPGK S290C

Figure D.7: 1) Left Column: smoothed and normalized optical density spectra OD′(λ) of A-slPGK (green) and unbound Al647-COOH (red). 2) Right Column: corrected and normalized (continuos line), and transmitted (dashed line) fluorescence emission spectra of A-slPGK (green) and unbound Al647-COOH (red). The total optical transmission function is also shown (black dashed line).

D.6 Measurement of the Detection Efficiency Ratio: g'

The measurement of the detection efficiency ratio g' was performed in two steps. In the first step, the reference value g'_R was determined with a smFRET calibration measurements of the two double-labeled DNA samples ("10bp" and "17bp"). In addition, the value of β_R was also measured on the same day. In the second step, the detection efficiency ratio g' was corrected for daily fluctuations every time a smFRET measurements was performed. The ensemble donor lifetimes of the double-labeled samples were also measured in order to obtain the width σ_{DA} of the gaussian distribution of inter-dye distances $p_G(r)$ induced by the linker dynamics. In fact, this value is needed to calculate the "corrected" static line.

D.6.1 eTCSPC Measurements

The donor lifetime of the double-labeled DNA samples was measured as described in Sec.6.3 and was fitted with the $F'_{DA}(t)$ model function (Eq.6.10) which considers a gaussian distribution $p_G(r)$ of inter-dye distances. Here, the background contribution was set to zero ($a_{bg} = 0$) because the Annealing buffer has a low background fluorescence. This means that the fitting parameters are the average inter-dye distance $\langle r_{DA} \rangle$, the width of the gaussian distribution of distances σ_{DA}, the fraction of the donor only species x_{D0}, and the amplitude at time zero a_D. The decay times ($\tau_{D0,1}$ and $\tau_{D0,2}$) and the fraction x_1 obtained from the donor only samples (Tab.9.3) were fixed during the fit. A Förster radius of 51.1 Å (Tab.9.5) was used to calculate the inter-dye distances $\langle r_{DA} \rangle$. The fitted decays and the fitting results are displayed in Fig.D.8 and in Tab.D.7. The width $\sigma_{DA} = 6$ Å, which is constant for the two samples, was used to calculate the "corrected" static line (Sec.D.6.2).

Table D.7: Fitting parameters of the eTCSPC decay of dlDNA. $R_0 = 51.1$ Å.

sample	$\langle r_{DA} \rangle$	σ_{DA}	x_{D0}	χ^2_r
dlDNA 10bp	38.9 Å	6.0 Å	0.35	1.015
dlDNA 17bp	55.2 Å	6.0 Å	0.15	1.044

D.6.2 smFRET Calibration Measurement with dlDNA

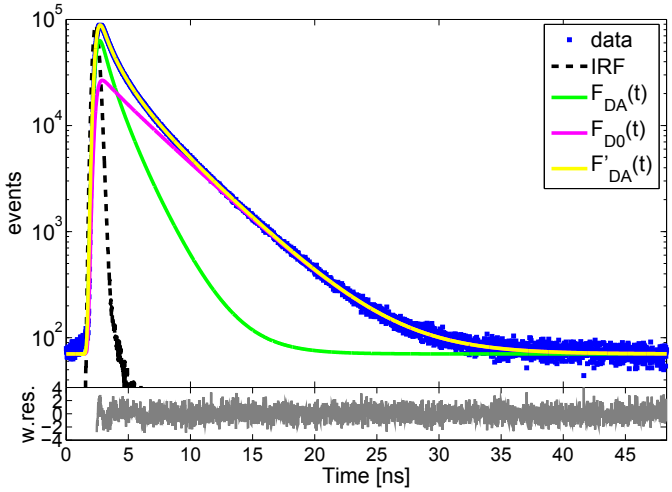
The smFRET calibration experiments were performed with the confocal microscope on the two double-labeled DNA samples ("10bp" and "17bp"). The excitation intensities (I_A and I_D) were set to the optimal values reported in Tab.4.2, and for both samples a set of four measurements of 1 hour was conducted. The same set-up used for the smFRET measurements with PGK

was calibrated. The concentration of the acceptor was set to a value equivalent to an apparent average number of molecules inside the confocal volume equal to ~ 0.03. Annealing buffer with $0.001\,\%$ of Tween®20 was used. The cover slides were cleaned for ten minutes inside a HPLC ethanol bath, rinsed with water, and dried with nitrogen. The data were analyzed with the already mentioned MATLAB scripts (Sec.7.5) in order to get the burst coordinates. Here, the following thresholds were used: $T_{IPD} = 160\,\mu s$ and $N_T = 30$. The 2D-plots were then built up with the parameters derived in Chap.9 for dlDNA. The static line was corrected as described in Sec.8.6 for the linker dynamics by using the gaussian width σ_{DA} obtained in the previous section and the donor only intensity averaged lifetime $\langle\tau_{D0}\rangle_f$ reported in Tab.9.3. The transformation used for the correction of the static line is displayed in Fig.D.9.

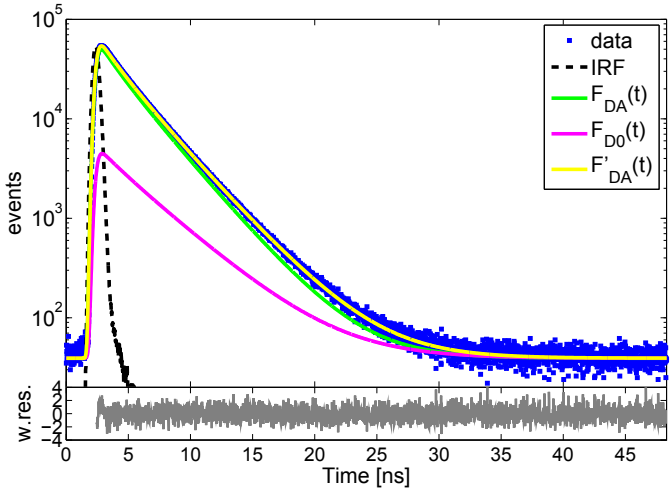
D.6.3 Daily Correction

In order to measure the ratio $\beta = \frac{f_A}{f_D}$ between the fluorescence detection rates f_A and f_D, a set of eight FCS measurements of 1 min was performed on the confocal microscope by exciting with the blue laser (470 nm) at 2000 a.u. the dye Rh101 dissolved in ethanol. Between two measurements the focus was moved in order to mimic the condition of the smFRET experiments. The ratio β was then calculated for each measurements, and the arithmetic mean $\bar{\beta}$ was calculated in order to reduce the uncertainities. Before performing these measurements, the absorption spectra of the dye was checked in order to rule out possible spectral shifts. In fact, a shift of the spectrum would influence β like a change of the detection efficiency ratio g' (see Eq.4.2).

D.6.4 Figures



(a) dlDNA 10bp (Fixed parameters: $\tau_{D0,1} = 1.42\,ns$, $\tau_{D0,2} = 4.03\,ns$, $x_1 = 0.12$ see Tab.9.3).



(b) dlDNA 17bp (Fixed parameters: $\tau_{D0,1} = 1.57\,ns$, $\tau_{D0,2} = 4.06\,ns$, $x_1 = 0.15$ see Tab.9.3).

Figure D.8: Fitted eTCSPC histogram of double-labeled DNA. The fitting parameters are reported in Tab.D.7.

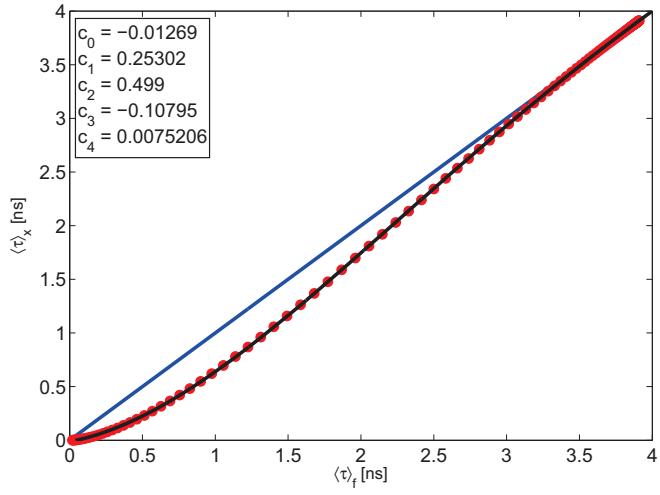


Figure D.9: The black line represents the polynomial empirical transformation used to correct the static line for DNA linker dynamics. The fitted polynomial coefficient $\{c_i\}$ are shown in the inset. The following parameters were used to calculate the set of $\langle\tau\rangle_x$ and $\langle\tau\rangle_f$ values (red circles): $\sigma_{DA} = 6$ Å and $\langle\tau_{D0}\rangle_f = 3.91$ ns.

Appendix E

PGK Interdomain Dynamics

E.1 Preparation of the Substrate's Stock Solutions

The substrates stock solutions were prepared as follows having in mind two main concerns: purity and pH. Each substrate purchased from Sigma®was dissolved in 5 ml of the FRET buffer. Here, 200 mg of 3-PG, and $\sim$ 160 mg of Mg_2ATP and KAPD were used. Afterwards, the solutions were first vortexed, and then filtered with 0.2 μm pore ACRODISC®syringe filters. Later on, the pH of the solutions were adjusted to pH 7.45 using a 2 M stock solution of ultrapure NaOH. Here, NaOH must be added in small volumes while stirring the solution in order to avoid overshooting of the pH which would degradate the substrates, especially ATP and ADP. The final concentration of 3-PG is circa 164 mM. On the other hand, the concentrations of the adenosine mojeties (ATP and ADP) were measured on the absorption spectrometer after a 1000x dilution achieved in two steps. In fact, the extinction coefficient of these species is $\epsilon = 15400\,M^{-1}cm^{-1}$ [Sigma®Product Information]. Only freshly prepared solutions were used for the measurements in order to avoid degradation of the substrates.

E.2 Fit of the eTCSPC Histograms

The donor intensity decays of dlPGK, which were obtained from the ensemble lifetime measurement carried out with and without substrates, were fitted in MATLAB with the model function $F'_{DA}(t)$ described in Sec.6.3.2. Here, the donor only lifetime components ($\{x_{D0,i}\}$ and $\{\tau_{D0,i}\}$) were fixed during the fit. These parameters are reported in Tab.E.1 and were estimated with Eq.9.1 by using the values measured with the single-labeled samples. A Förster radius of 51.0 Å was used to calculate the inter-dye distance. The fitting parameters are reported in Tab.E.2. The gaussian distribution of distances $p_G(r_{DA})$ obtained from the fits are displayed in Fig.E.1, and the fitted curves are shown in Fig.E.2.

Table E.1: Donor only decay times $\{\tau_{D0,i}\}$ and amplitude fractions $\{x_{D0,i}\}$ fixed during the fitting procedure.

substrate	$\tau_{D0,1}$ [ns]	$\tau_{D0,2}$ [ns]	$x_{D0,1}$	$x_{D0,2}$
none	1.65	3.98	0.17	0.83
Mg_2ATP	1.46	3.88	0.17	0.83
KAPD	1.52	3.90	0.16	0.84
3-PG	1.93	4.05	0.15	0.85
3-PG*KAPD	1.70	3.94	0.17	0.83

Table E.2: Fitting parameters obtained from the fit of the donor intensity decays using the $F'_{DA}(t)$ model function reported in Sec.6.3.2. The Förster radius is 51.0 Å.

substrate	$\langle r_{DA} \rangle$ [Å]	σ_{DA} [Å]	x_{D0} (%)	x_{bg} (%)	τ_{bg} [ns]	χ^2_r
none	45.2	8.8	50	0.02	29.3	1.286
Mg_2ATP	48.4	8.1	63	0.12	16.6	1.260
KADP	47.2	8.2	58	0.06	18.4	1.104
3-PG	45.8	8.7	56	0.04	17.9	1.225
3-PG*KADP	46.2	8.8	58	0.17	19.5	1.160

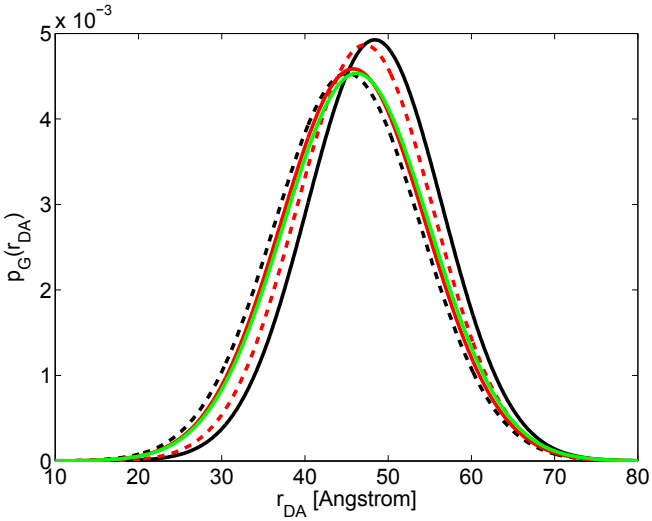
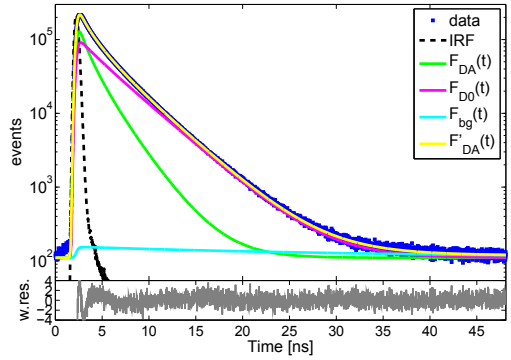
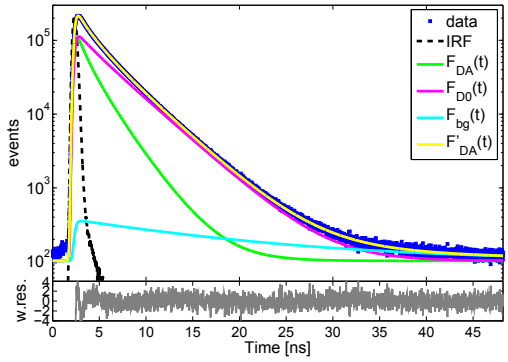


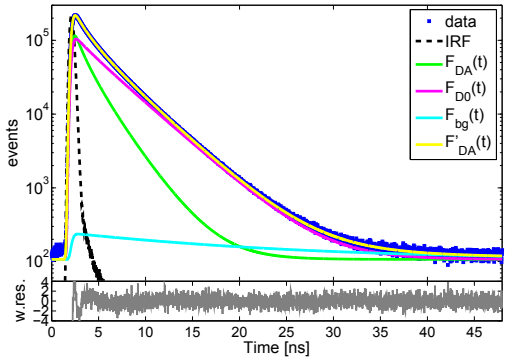
Figure E.1: Gaussian distributions $p_G(r_{DA})$ of the inter-dye distances r_{DA} normalized with respect to the area. The curves were obtained by fitting the intensity decays of the donor with 3-PG (red continuos), Mg_2ATP (black continuos), KAPD (red dashed), 3-PG*KAPD (green) and without substrate (dashed black). A Förster radius of 51.0 Å was used to calculate the distances.



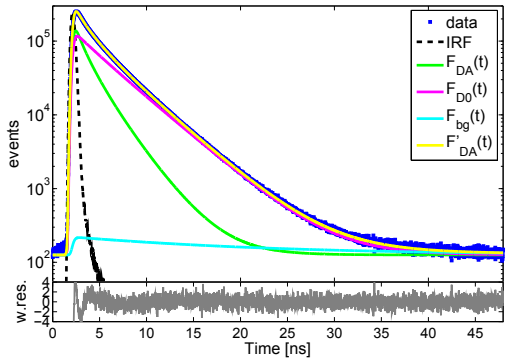
(a) none



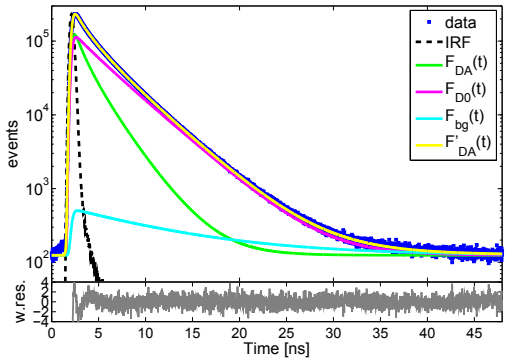
(b) Mg_2ATP



(c) KADP



(d) 3-PG



(e) 3-PG*KADP

Figure E.2: Fitted decays of the donor bound to dIPGK measured with and without substrates.

E.3 smFRET Measurements

In this section, some precautions applied to perform the smFRET measurements on dlPGK are first described. Then, the procedures followed to analyze the data with the 1D efficiency histograms and the 2D-plots are illustrated.

E.3.1 smFRET Measurements with dlPGK

The smFRET experiments were conducted following the general strategy described in Sec.7.6. In addition, before performing the measurements, the freshly defrosted dlPGK aliquots were tested with FCS in order to assure the goodness of the sample, and the daily correction of the detection efficiency ratio g' was determined as described in Sec.9.7.3. Diffusion coefficients fluctuating in a range of $(53-56)\,\frac{\mu m^2}{s}$ were observed. These values are lower than the ones ($D = 68\,\frac{\mu m^2}{s}$) calculated from the hydrodynamic radius measured by DLS (see Sec.3.1.3). In fact, as shown in the main text (see Sec.10.5), at low protein concentrations oligomers are present which disappears when the concentration increases. Furthermore, the dlPGK aliquots were kept on ice during all day. Only before each of the one hour smFRET measurement the stock was diluted in FRET buffer to the optimal single-molecule concentration ($N_{FCS} \sim 0.03$), after addition of $0.004\,\%$ of Tween®20 and the desired substrate. In fact, partial unfolding of PGK, which appears with an increase of the population at low FRET ($E_2 \sim 0.15$), was observed for samples stored at room temperature for many hours. For this reason, the $5-6$ hours lasting measurements performed overnight at the end of the day were always checked by looking at the 2D-plots. If unfolding was detected, the corresponding files were discarded from the analysis.

E.3.2 Fit of the 1D Efficiency Histograms

The 1D efficiency histograms of dlPGK in the bound and the unbound states (see Fig.10.3 and Fig.10.2a) were constructed with the set of parameteres derived in Chap.9 and summarized in Tab.E.3. The efficiency histograms were fitted in MATLAB with a three gaussian model function defined in the main text by Eq.8.11 for $m = 3$. Here, the weighted least squares estimator (LSE) was mimimized with MINUIT using a poissonian distribution to evaluate the errors on each data point ($\sigma_i = \frac{1}{\sqrt{N_i}}$) [54]. The unbound state of PGK was fitted at first, and the bound states were analyzed thereafter by fixing the value of $\langle E_2 \rangle = 0.15$ to the value obtained for the unbound state. The fitting parameters are reported in Tab.E.4 and the fitted histograms in the main text (Sec.10.4).

Table E.3: Set of parameters employed for the construction of the 1D efficiency histograms. Here, α and γ' were the same for the whole set of conditions. See Chap.9 for further details about the measurements of the parameters.

substrate	α	γ'	g'
none	0.039	0.67	0.76
Mg_2ATP	/	/	0.78
KAPD	/	/	0.70
3-PG	/	/	0.76
3-PG*KAPD	/	/	0.77

Table E.4: Set of parameters obtained from the fit of the 1D efficiency histograms of dlPGK in the bound and the unbound states. Here, for each gaussian component the mean $\langle E_i \rangle$, the variance σ_i^2 and the relative population x_i are reported. The subscripts represents the partially closed population ($i = 1$), the open population ($i = 2$) and the mixed population ($i = 12$). The mean efficiency $\langle E_2 \rangle$ of the open population was fixed to the value obtained for the unbound state.

substrate	$\langle E_1 \rangle$	$\langle E_{12} \rangle$	$\langle E_2 \rangle$	σ_1^2	σ_{12}^2	σ_2^2	x_1	x_{12}	x_2	χ_r^2
none	0.80	0.57	0.15	0.009	0.025	0.024	0.48	0.33	0.20	2.091
Mg_2ATP	0.70	0.41	/	0.013	0.011	0.018	0.78	0.11	0.11	1.497
KAPD	0.75	0.51	/	0.009	0.031	0.018	0.65	0.26	0.09	1.547
3-PG	0.77	0.54	/	0.008	0.035	0.025	0.65	0.28	0.05	1.527
3-PG*KAPD	0.75	0.55	/	0.009	0.025	0.024	0.66	0.29	0.06	1.435

E.3.3 Correction of the Apparent Inter-Dye Distances

The apparent FRET averaged inter-dye distances $\langle \tilde{r} \rangle$, which were calculated with Eq.8.2 employing the mean efficiencies $\langle E \rangle$ obtained from the fit of the efficiencies histograms, are biased towards longer distances by the acceptor quantum yield distribution $p(\phi_A)$. In order to correct for this effect, the following equation was used [68]:

$$\langle r_{DA} \rangle_E = \frac{\langle \tilde{r} \rangle}{\langle \phi_A \rangle^{\frac{1}{6}} \cdot \langle \phi_A^{-\frac{1}{6}} \rangle} = \frac{\langle \tilde{r} \rangle}{\beta} \tag{E.1}$$

Here, the denominator β can be calculated if the decay times $\{\tau_i\}$ and the amplitude fractions $\{x_i\}$, obtained from the fit of the eTCSP histogram of the directly excited acceptor, are known. In fact, the average quantum yield $\langle \phi_A \rangle^{\frac{1}{6}}$ is defined by the following expression derived from Eq.9.7:

$$\langle \phi_U \rangle^{\frac{1}{6}} = \left[\frac{\langle \tau_U \rangle_x}{\langle \tau_R \rangle_x} \cdot \langle \phi_R \rangle \right]^{\frac{1}{6}} \tag{E.2}$$

The second term at the denominator $\langle \phi_U^{-\frac{1}{6}} \rangle$ is derived through the definition of ensemble average $\langle \phi \rangle$ for a discrete number of states (see Eq.D.1):

$$\langle \phi_U^{-\frac{1}{6}} \rangle = \sum_i x_{i,U} \cdot \phi_{i,U}^{-\frac{1}{6}} = \sum_i x_{i,U} \cdot \left[\frac{\tau_{i,U}}{\langle \tau_R \rangle_x} \cdot \langle \phi_R \rangle \right]^{-\frac{1}{6}} \tag{E.3}$$

Here, the analogy with Eq.9.6 was applied in order to derive an expression for the quantum yield component $\phi_{i,U} = \frac{\tau_{i,U}}{\langle \tau_R \rangle_x} \cdot \langle \phi_R \rangle$. Therefore, the FRET averaged distances $\langle r_{DA} \rangle_E$ are easily calculated with Eq.E.1 employing the mean apparent FRET averaged distances $\langle \tilde{r} \rangle$ of the individual FRET populations. The results are reported in Tab.E.5 and show a nearly constant deviation of $\sim 1.7\,\%$ due to the acceptor quantum yield distribution.

Table E.5: The mean apparent FRET averaged distances $\langle \tilde{r}_i \rangle$ and the FRET averaged distances $\langle r_{DA} \rangle_{E_i}$ are reported for each population and for the different substrates, together with the correction factor $\beta = \langle \phi_A \rangle^{\frac{1}{6}} \cdot \langle \phi_A^{-\frac{1}{6}} \rangle$ calculated with Eq.E.2 and Eq.E.3. The following lifetime and quantum yield were used as reference values: $\langle \tau_R \rangle_x = (0.36 \cdot 1.19\,\text{ns}) + (0.64 \cdot 1.49\,\text{ns}) = 1.38\,\text{ns}$ (Tab.D.1), $\langle \phi_R \rangle = 0.58$ (Tab.D.4).

substrate	β	$\langle \tilde{r}_1 \rangle$	$\langle r_{DA} \rangle_{E_1}$	$\langle \tilde{r}_{12} \rangle$	$\langle r_{DA} \rangle_{E_{12}}$	$\langle \tilde{r}_2 \rangle$	$\langle r_{DA} \rangle_{E_2}$
none	1.0171	40.4	39.7	48.8	48.0	68.3	67.2
Mg_2ATP	1.0167	44.3	43.6	54.3	53.4	/	/
KAPD	1.0179	42.4	41.7	50.8	49.9	/	/
3-PG	1.0164	41.6	40.9	49.7	48.9	/	/
3-PG*KAPD	1.0165	42.5	41.8	49.3	48.5	/	/

E.3.4 Calculation of σ^2_{SN} and $\sigma^2_{\phi_A}$

The shot noise variance σ^2_{SN} and the variance due to fluctuations of the acceptor quantum yield $\sigma^2_{\phi_A}$ were calculated as described in Appx.C.3 for each population observed in the 1D efficiency histograms. The mean FRET efficiencies $\langle E_i \rangle$ and the FRET averaged distances $\langle r_{DA} \rangle_{E_i}$ reported in Tab.E.4 and Tab.E.5 were used for the computation. The results are reported in Tab.E.6, Tab.E.7, and Tab.E.8, where also the total variance $\sigma^2_{tot} = \sigma^2_{SN} + \sigma^2_{\phi_A}$ and the apparent variance of the inter-dye distance due to the acceptor quantum yield distribution $\sigma^2_{\tilde{r}}$ (see Appx.C.3.2) are shown. The distributions p(N) of the total number of photons detected in one fluorescence burst, which were used to calculate the ensemble average $\langle N \rangle$ with Eq.C.10, are displayed in Fig.E.3.

Table E.6: Calculated variances of the partially closed populations: i = 1.

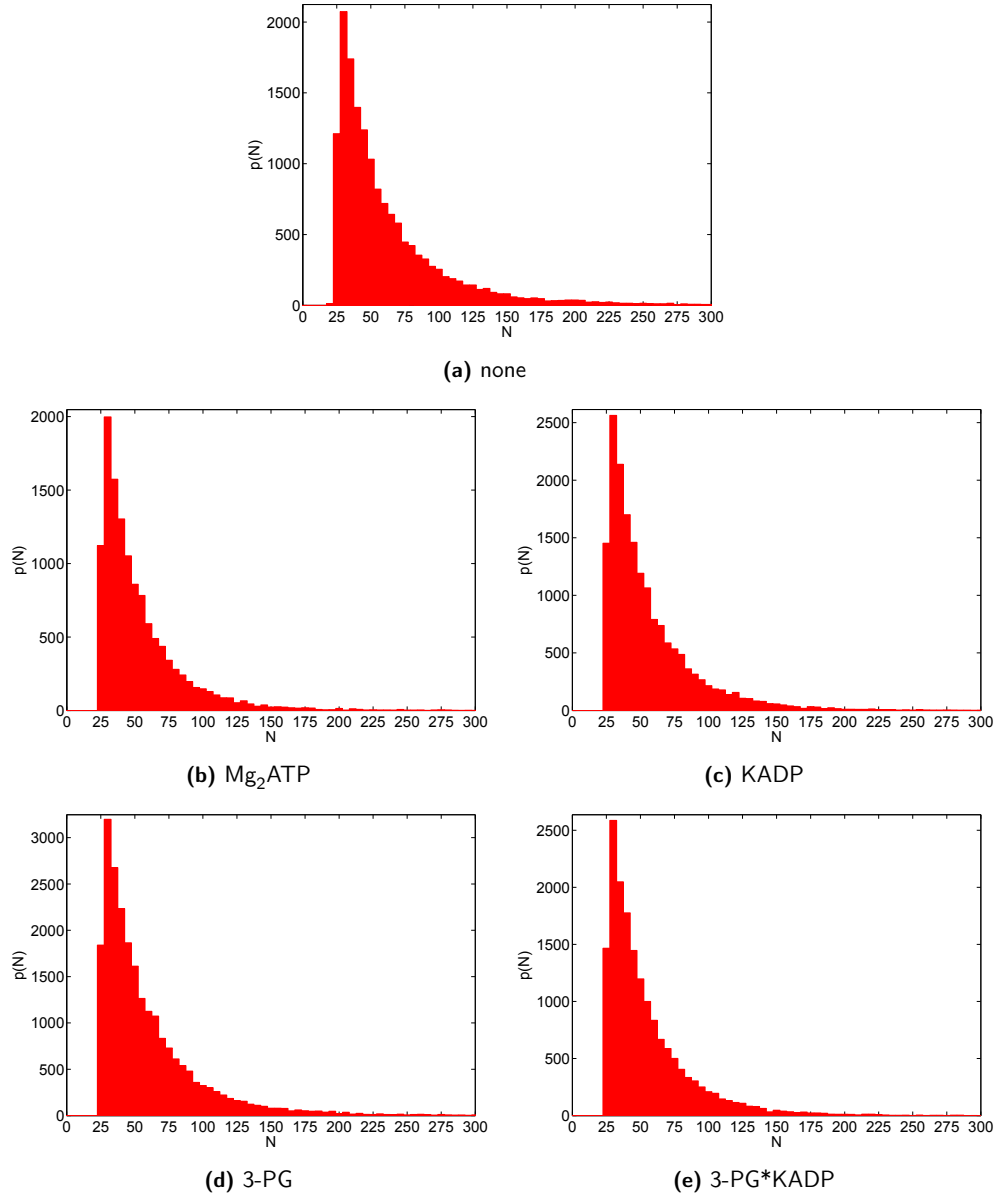
substrate	$\langle N \rangle$	$\sigma^2_{SN,1}$	$\langle r_{DA} \rangle_{E_1}$ [Å]	$\sigma^2_{\tilde{r}_1}$ [Å]	$\sigma^2_{\phi_{A,1}}$	$\sigma^2_{tot,1}$
none	70.6	0.0023	39.7	3.1	0.0054	0.0075
Mg_2ATP	61.4	0.0034	43.6	3.3	0.0086	0.0120
KAPD	70.2	0.0027	41.7	3.3	0.0075	0.0101
3-PG	61.5	0.0029	40.9	3.1	0.0061	0.0090
3-PG*KAPD	68.3	0.0027	41.8	3.2	0.0069	0.0096

Table E.7: Calculated variances of the mixed populations: i = 12.

substrate	$\langle N \rangle$	$\sigma^2_{SN,12}$	$\langle r_{DA} \rangle_{E_{12}}$ [Å]	$\sigma^2_{\tilde{r}_{12}}$ [Å]	$\sigma^2_{\phi_{A,12}}$	$\sigma^2_{tot,12}$
none	70.6	0.0035	48.0	3.7	0.0112	0.0157
Mg_2ATP	61.4	0.0034	53.4	4.0	0.0110	0.0150
KAPD	70.2	0.0036	49.9	4.0	0.0132	0.0168
3-PG	61.5	0.0040	48.9	3.7	0.0040	0.0161
3-PG*KAPD	68.3	0.0036	48.5	3.7	0.0118	0.0154

Table E.8: Calculated variances of the open populations: i = 2.

substrate	$\langle N \rangle$	$\sigma^2_{SN,2}$	$\langle r_{DA} \rangle_{E_2}$ [Å]	$\sigma^2_{\tilde{r}_2}$ [Å]	$\sigma^2_{\phi_{A,2}}$	$\sigma^2_{tot,2}$
none	70.6	0.0018	67.2	5.2	0.0035	0.0053
Mg_2ATP	61.4	0.0021	/	5.1	0.0037	0.0055
KAPD	70.2	0.0018	/	5.3	0.0037	0.0055
3-PG	61.5	0.0020	/	5.1	0.0033	0.0054
3-PG*KAPD	68.3	0.0018	/	5.1	0.0033	0.0051



(a) none

(b) Mg_2ATP

(c) KADP

(d) 3-PG

(e) 3-PG*KADP

Figure E.3: Experimental distributions p(N) of the single-burst total fluorescence intensity $N = F_A + F_D$ detected with the smFRET measurements.

E.3.5 "Corrected" Static Line, Dynamic Line and 2D-Plots

The 2D-plots shown in Fig.10.2b and Fig.10.3 in the main text were calculated with the set of parameters reported in Tab.E.3. The "corrected" static lines and the 2-state dynamic lines were calculated as described in Appx.C.4. In order to do so, the parameters reported in Tab.E.9 and the polynomial coefficients $\{c_i\}$ given in Tab.E.10 and Tab.E.11, which were obtained from the fit of the $\langle\tau\rangle_x$ and $\langle\tau\rangle_f$ values displayed in Fig.E.4, were used. In order to calculate the gaussian distributions $p_{2G}(r_{DA})$ (Eq.C.19) needed to obtain the dynamic lines, the mean efficiencies $\langle E_1\rangle$ and $\langle E_2\rangle$ shown in Tab.E.4 were employed.

Table E.9: Set of parameters used for the calculation of the "corrected" static lines and the 2-state dynamic lines.

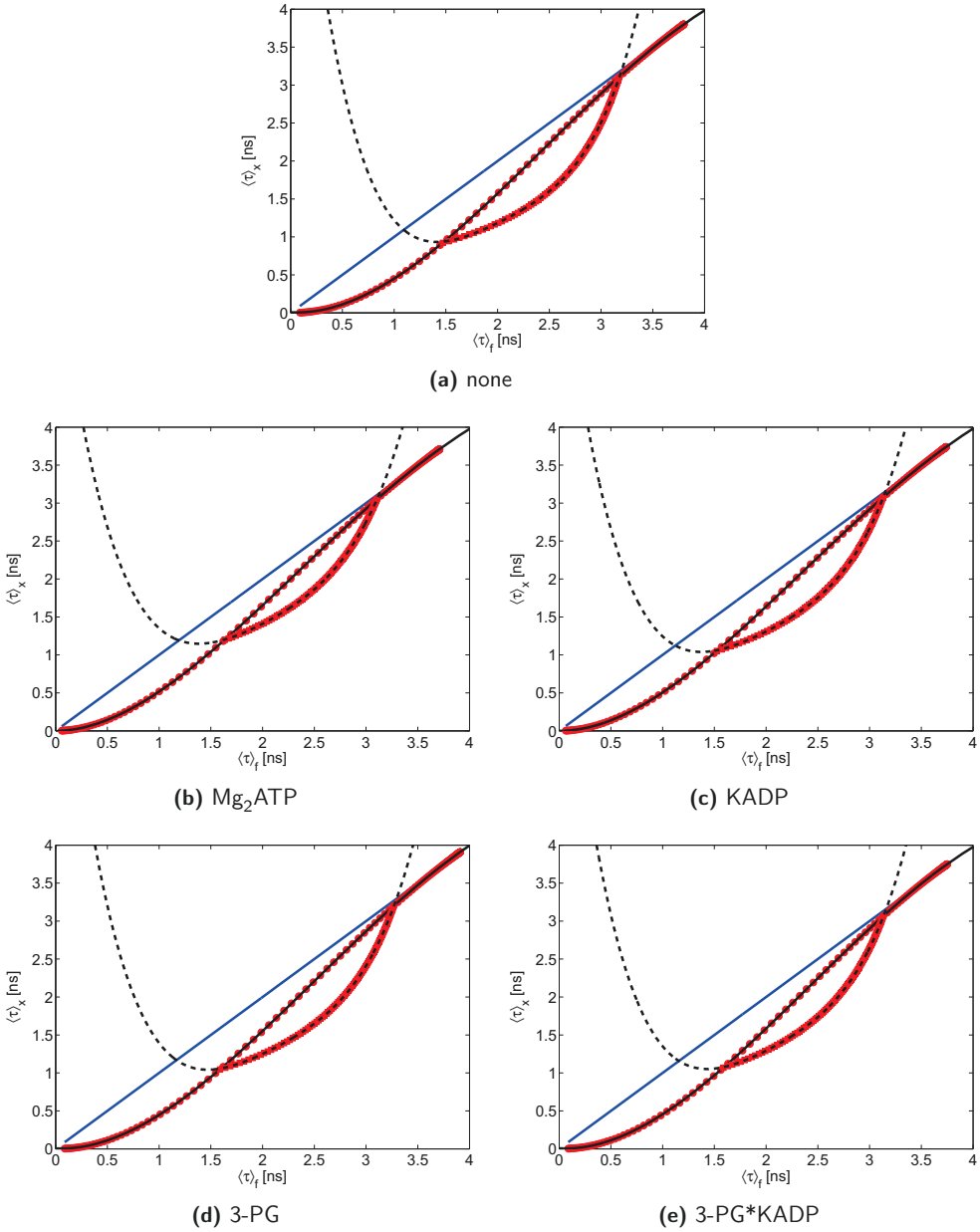
substrate	R_0 [Å]	$\langle\tau_{D0}\rangle_f$ [ns]	σ_{DA} [Å]
none	51.0	3.80	8.8
Mg_2ATP	/	3.71	8.1
KAPD	/	3.74	8.2
3-PG	/	3.91	8.7
3-PG*KAPD	/	3.75	8.8

Table E.10: Polynomial coefficients $\{c_i\}$ used to calculate the "corrected" static lines.

substrate	c_0	c_1	c_2	c_3	c_4
none	0.0073569	-0.074328	0.61152	-0.097116	0.002751
Mg_2ATP	-0.0019235	0.014638	0.61101	-0.11302	0.0053779
KAPD	-0.00088805	0.0015603	0.60903	-0.10995	0.0049474
3-PG	0.0059168	-0.062056	0.59302	-0.093456	0.0028539
3-PG*KAPD	0.0072601	-0.074328	0.61968	-0.099723	0.0028625

Table E.11: Polynomial coefficients $\{c_i\}$ used to calculate the dynamic lines.

substrate	c_0	c_1	c_2	c_3	c_4
none	7.9472	-14.4427	10.7679	-3.4843	0.43274
Mg_2ATP	6.3356	-10.625	7.8222	-2.4921	0.31084
KADP	6.66169	-11.6018	8.6937	-2.8195	0.35479
3-PG	7.9087	-13.5363	9.654	-2.9902	0.35754
3-PG*KADP	7.7448	-13.5417	9.9167	-3.1584	0.38954



(a) none

(b) Mg_2ATP

(c) KADP

(d) 3-PG

(e) 3-PG*KADP

Figure E.4: Polynomial fits of the set of $\langle\tau\rangle_x$ and $\langle\tau\rangle_f$ values calculated as described in Appx.C.4 and used to determine the "corrected" static lines (circles) and the two states dynamic lines (squares).

E.3.6 Conformational Variance of the Open State

The amplitude of the space sampled by the domains of PGK in the unliganded open state was determined with Eq.C.14. In order to do so, the conformational variance $\sigma^2_{r,2} = \sigma^2_2 - \sigma^2_{tot,2}$ (see Eq.8.10) was calculated as the difference between the measured variance σ^2_2 (Tab.E.4) and the total variance $\sigma^2_{tot,2}$ (Tab.E.8), that is the sum of $\sigma^2_{SN,2}$ and $\sigma^2_{\phi_{A,2}}$. Then, a gaussian distribution of distances was assumed $p(r) \sim p_G(r)$ (Eq.C.18), in order to rewrite Eq.C.14 in an empirical form:

$$\sigma^2_r \simeq \int [E(r) - \langle E \rangle]^2 \cdot p_G(r)\, dr \tag{E.4}$$

Here, E(r) is defined by Eq.8.2. Therefore, after setting the mean of the gaussian $\langle r_{DA} \rangle$, the mean efficiency $\langle E \rangle$, and the Förster radius R_0 to the measured values ($\langle E_2 \rangle$, $\langle r_{DA} \rangle_{E,2}$ and R_0), the integral is a function solely of the standard deviation σ_{DA}. Thus, the solution of Eq.E.4 was obtained by tuning σ_{DA} until the calculated conformational variance σ^2_r reached the measured value of $\sigma^2_{r,2} = 0.0187$. The graphical solution $\sigma_{DA} \sim 8.9$ Å is given in Fig.E.5 by the intersection between the red line and the black dashed line. Therefore, the average positions of the dye explore a space with an extension of ~ 17.8 Å because of the rigid motions of the domains.

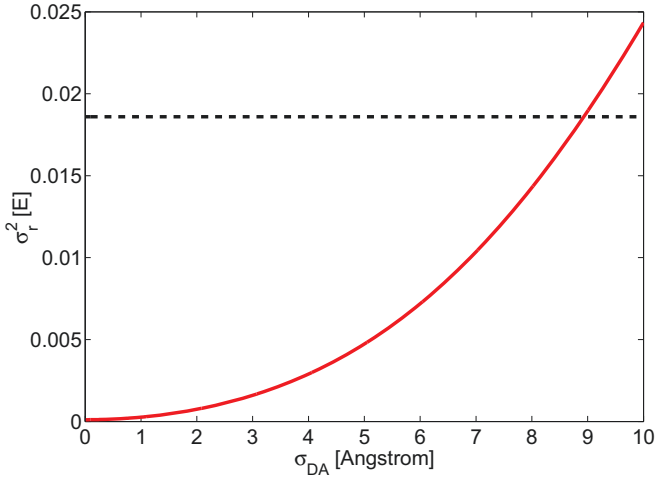


Figure E.5: The values of σ^2_r calculated with Eq.E.4 are shown as a function of the standard deviation σ_{DA} of the gaussian distribution of distances $p_G(r)$ (red line). The black dashed line indicates the measured conformational variance $\sigma^2_{r,2} = 0.0187$ (see Tab.E.8). The following parameters were used to calculate the integral with Eq.E.4: $\langle E_2 \rangle = 0.15$, $\langle r_{DA} \rangle_{E,2} = 67.2$ Å and $R_0 = 51.0$ Å.

E.3.7 The State Occupation Probabilities at the Equilibrium

The set of occupation probabilities $\{p_i\}$ under a given set of external conditions defines the most probable distribution of states of the system [81]. For a protein at the equilibrium, the probability p_i to be in a conformational state-i is determined by the Boltzmann distribution:

$$p_i = \frac{e^{-\beta \cdot \Delta G_i}}{Z} \tag{E.5}$$

Where, $e^{-\beta \cdot \Delta G_i}$ is the statistical weight or Boltzmann factor of state-i defined in terms of the relative Gibbs free energy ($\Delta G_i = \Delta H_i - T\Delta S_i$), $Z = \sum_i e^{-\beta \cdot \Delta G_i}$ is the conformational partition function, which defines all the conformations accessible by the protein, and $\beta = \frac{1}{k_B T} = 2.45 \times 10^{20}\ \frac{1}{J}$. The previous equation shows that the set of conformations $\{i\}$ adopted by the protein at the equilibrium depends on the relative free energies of each state ΔG_i, that are a function of the external conditions. Therefore, an external perturbation will induce a redistribution of the occupation probabilities $\{p_i\}$ with a consequent rearrangments of the protein conformations. The concepts of energy landscape and dynamic energy landscapes discussed in the main text (Sec.10.4.3) arise from the statistical description of the conformational ensemble of a protein given above. Here, the Gibbs free energy is normally represented as a function of the reaction coordinate, that is a dynamically relevant measure of the progress of the reaction [109]. In the present work, the inter-dye distance r_{DA} was choosen in order to depict the conformational motions of the protein domains.

To be more practical, the previous equation (Eq.E.5) shows that the relative Gibbs free energies ΔG_i for each state-i can be calculated if the whole set of occupation probabilities $\{p_i\}$ is known. Since the occupation probability by definition is a function of the number of molecules N_i in a specific state-i, the relative free energies ΔG_i are easily obtained from smFRET experiments where the molecules in one state-i are inherently counted:

$$p_i = \frac{N_i}{\sum_i N_i} \tag{E.6}$$

For a two state model system ($i = 1, 2$), where the two states can correspond to a couple of individual conformations or to sub-ensembles of energetically related conformers, Eq.E.5 simplifies to:

$$p_i = \frac{1}{1 + e^{-\beta \Delta G_{ij}}} \tag{E.7}$$

Where, $\Delta G_{ij} = \Delta G_j - \Delta G_i$ is the relative Gibbs free energy between the two states that can be calculated by rearranging Eq.E.7 and considering that $\sum_i p_i = 1$:

$$\Delta G_{ij} = -\frac{\ln\left(\frac{p_j}{p_i}\right)}{\beta} \tag{E.8}$$

This equation shows that the relative populations of the two states i and j at the equilibrium depend only on their relative free energy ΔG_{ij} and on the temperature. The interesting physical information is enclosed within the energetic term, which depends on the conformations assumed by the protein under the conditions imposed by the surrounding environment (T, P, pH, crowding, ligands, etc.). A change of each of the external parameter can tune the imbalance between the occupation probabilities of the two states influencing the protein behavior.
The equation above (Eq.E.8) was used to calculate the relative free energies (ΔG_{12} and ΔG_{43}) between the open and the partially closed conformations of the domains in the ligand-free and ligand-bound states of PGK (Tab.10.2). Additionally, it can be applied to determine even more precise ΔG_{ij} values by measuring the occupation probabilities at different temperatures. With the definition given in Eq.E.8, the energies are in units of $\left[\frac{J}{molecule}\right]$. Therefore, in order to obtain the desired units of $\left[\frac{kcal}{mol}\right]$, the calculated energies must be multiplied by the Avogadro's number $N_a = 6.0221 \times 10^{23}\,\frac{1}{mol}$ and by the conversion factor from Joules to calories ($0.239\,\frac{cal}{J}$).

E.3.8 The Kinetic Equilibrium Constants

Further information about the two state system at the equilibrium can be extracted employing a chemical kinetics approach. The following two state kinetic sheme was used, and exclusively unimolecular reactions were considered:

$$[i] \underset{k_b}{\overset{k_f}{\rightleftharpoons}} [j] \tag{E.9}$$

Here, k_f and k_b are the microscopic rate constants for the forward and the backward reactions, respectively. Accordingly to the transition-state theory [110] the microscopic rates k are defined as:

$$k = \frac{1}{\beta h} \cdot e^{-\beta \Delta G^{\ddagger}} \tag{E.10}$$

Where, h is the Planck's constant ($6.626 \times 10^{34}\,J \cdot s$) and $\Delta G^{\ddagger}$ is the activation energy to cross the barrier separating the two states i and j. Therefore, the transition rate is a function of the barrier height $\Delta G^{\ddagger}$, the higher the barrier the lower the probability for a transition to take place. An additional quantity, the kinetic equilibrium constant K_{ij}, can be defined from the microscopic rate constants in order to describe the equilibrium between the two populations i

and j:

$$K_{ij} = \frac{k_f}{k_b} = \frac{[j]}{[i]} \tag{E.11}$$

Here, [i] and [j] are the species concentrations. From this equation, a further relation is derived using the definition of species concentration $[i] = \frac{N_i}{V}$, multipling by $\frac{\sum_i N}{\sum_i N} = 1$, and considering Eq.E.6:

$$K_{ij} = \frac{p_j}{p_i} \tag{E.12}$$

This expression was used to compute the equilibrium constant between the open $(i = 1, 4)$ and the partially closed $(j = 2, 3)$ populations of the PGK domains in the bound and unbound states from the ratios of the occupation probabilities of the two states ($\frac{p_2}{p_1}$ and $\frac{p_3}{p_4}$). In addition, together with Eq.E.11, it shows that the occupation probabilities of the two states depend on the ratio between the microscopic rates for the forward k_f and the backward k_b reactions. Finally, a substitution of Eq.E.10 in Eq.E.11 permits to relate the equilibrium constant K_{ij} to the relative Gibbs free energy between the two states ΔG_{ij}:

$$K_{ij} = e^{-\beta[\Delta G_i^\ddagger - \Delta G_j^\ddagger]} \tag{E.13}$$

In fact, the difference between the activation energies is exactly the relative energy between the two states i and j:

$$\Delta G_{ij} = \Delta G_i^\ddagger - \Delta G_j^\ddagger \tag{E.14}$$

Therefore, the equilibrium constant depends solely on the relative free energy between the two states ΔG_{ij} and not on the heights of the barriers ($\Delta G_i^\ddagger$ and $\Delta G_j^\ddagger$), that on long enough time interval influence only the transition probabilities but not the equilibrium distributions of states.

E.3.9 The Opening and Closing Activation Barriers

For the kinetic scheme represented in Eq.E.9, the characteristic interconversion time τ_c between the two states i and j is the sum of the the microscopic rate constants for the forward k_f and the backward reactions k_b [111]:

$$\frac{1}{\tau_c} = k_f + k_b \tag{E.15}$$

This expression can be rearranged by expliciting k_f and k_b as a function of the activation energies $\Delta G_i^\ddagger$ and $\Delta G_j^\ddagger$ with Eq.E.10 that gives:

$$\Delta G_i^\ddagger + \Delta G_j^\ddagger = -\frac{\ln\left(\frac{1}{\tau_c}\right) + \ln(\beta h)}{\beta} \tag{E.16}$$

Since the two energy terms are both unknown, a further relation must be derived to have a working equation. Thence, an explicit relation between the two activation energies is derived using Eq.E.8 and considering that $\Delta G_{ij} = \Delta G_i^{\ddagger} - \Delta G_j^{\ddagger}$:

$$\Delta G_j^{\ddagger} = \Delta G_i^{\ddagger} - \frac{\ln\left(\frac{p_j}{p_i}\right)}{\beta} \tag{E.17}$$

This relation is substituted in Eq.E.16 to give the desired equation where only measurable quantities appear:

$$\Delta G_i^{\ddagger} = -\frac{\ln\left(\frac{p_j}{p_i}\right) + \ln\left(\frac{1}{\tau_c}\right) + \ln(\beta h)}{2\beta} \tag{E.18}$$

Therefore, this equation in combination with Eq.E.17 was used to calculate the activation energies reported in the main text (Tab.10.2) from the measured occupation probabilities (p_i and p_j) and the estimated relaxation correlation time τ_c. The barriers correspond to the transitions from the open to the closed states of PGK and vice versa.

E.4 Viscosity of the Substrates Solutions

The viscosities η of the solutions employed to perform the FCS, the smFRET, and the eFRET measurements with and without the substrates were measured with the rolling ball method. The following equation was used to determine the unknown viscosity η [mPa · s] from the measured falling time t_1 [s]:

$$\eta = K_1 \cdot (\rho_K - \rho_P) \cdot t_1 \tag{E.19}$$

Here, K_1 $\left[\frac{mPa \cdot cm^3}{g}\right]$ is a calibration constant dependent on the tilting angle θ° of the capillar with respect to the vertical position, ρ_K $\left[\frac{g}{cm^3}\right]$ is the density of the ball, and ρ_P $\left[\frac{g}{cm^3}\right]$ is the density of the solvent. A capillar with a diameter of 1.600 mm and a ball with a diameter of 1.500 mm and a density $\rho_K = 7.670\,\frac{g}{cm^3}$ were utilized. The densities of the solutions, as a first approximation, were set to the value of water $\rho_P = 0.99820\,\frac{g}{cm^3}$. In fact, at salt concentrations of a few tens of mM the variations with respect to pure water are neglegible. The measurements were repeated at two different tilting angles (50° and 70°) and for each angle the falling time was averaged over 7 runs in order to increase the precision. The relative error on the estimated times t_1 is $\pm 0.02\,\%$. The temperature of the solutions was set to 23 °C. The results are displayed in Tab.E.12 for each solution.

Table E.12: Measured falling times t_1 and viscosities η of the substrates solutions. The calibration parameters K_1 are provided by the manufacturer of the instrument as a function of the tilting angle θ°: $K_1^{50^\circ} = 0.00781$ and $K_1^{70^\circ} = 0.00960$.

substrate	θ°	K_1 $\left[\frac{mPa \cdot cm^2}{s}\right]$	t_1 [s]	η [mPa · s]
none	50°	0.00781	18.23	0.950
none	70°	0.00960	14.82	0.950
Mg_2ATP	50°	0.00781	18.67	0.973
Mg_2ATP	70°	0.00960	15.19	0.973
KAPD	50°	0.00781	18.62	0.970
KAPD	70°	0.00960	15.14	0.969
3-PG	50°	0.00781	18.97	0.988
3-PG	70°	0.00960	15.43	0.988
3-PG*KAPD	50°	0.00781	19.37	1.009
3-PG*KAPD	70°	0.00960	15.78	1.010

E.5 Refractive Index Mismatch Correction

Mismatches between the refractive index of the measuring solution with respect to the refractive index of water ($n_{H_2O} = 1.333$) can have a dramatic impact on the measured diffusion coefficient, as shown in Fig.E.6, where the apparent diffusion coefficient is calculated as a function of the refractive index [83]. In fact, the water-immersion objective is optimally corrected for imaging in water. Therefore, the curve displayed in Fig.E.6 was used to estimate the deviation $\delta D\,\%$ of the measured values $D^{23\,^\circ C}$ from the real diffusion coefficients $D_{corr}^{23\,^\circ C}$ of dlPGK in the bound and unbound states, the results are reported in Tab.E.13.

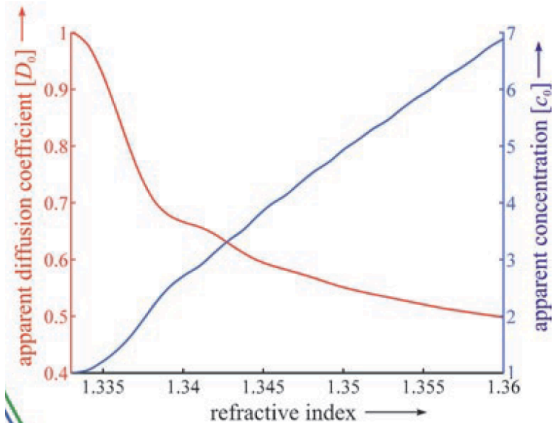


Figure E.6: Dependence of the apparent diffusion coefficient and concentration on the refractive index. See [83] for more details.

Table E.13: Measured $D^{23\,^\circ C}$ and corrected $D_{corr}^{23\,^\circ C}$ diffusion coefficients of dlPGK in the bound and unbound states.

substrate	n	$D^{23\,^\circ C}$ $\left[\frac{\mu m^2}{s}\right]$	$D_{corr}^{23\,^\circ C}$ $\left[\frac{\mu m^2}{s}\right]$	$\delta D\,(\%)$
none	1.3340	58	59.2	$\sim 2\,\%$
Mg_2ATP	1.3353	68	73.9	$\sim 8\,\%$
KAPD	1.3355	65	70.7	$\sim 8\,\%$
3-PG	1.3354	66	71.7	$\sim 8\,\%$
3-PG*KAPD	1.3369	66	77.6	$\sim 15\,\%$

E.6 Calculation of the Average Burst Duration

The average duration $\langle T \rangle$ of a set of the bursts, which were obtained from the smFRET measurements of ligand-free and liganded dlPGK, was calculated from the experimentally determined distributions $p(T) = \mathrm{hist}(T)$ without any assumption about their functional dependence:

$$\langle T \rangle = \frac{\int T \cdot p(T)\,dT}{\int p(T)\,dT} = \frac{\sum_i T \cdot \mathrm{hist}(T)}{\sum_i \mathrm{hist}(T)} \tag{E.20}$$

Here, the bursts were sorted out in two groups belonging to the open and the partially closed populations depending on their FRET efficiencies with respect to a threshold value N_{E_i}. The threshold was set in order to discard burst belonging to the mixed population (see Fig.E.6 and Fig.10.3). The results are reported in Tab.E.14.

Table E.14: Average burst durations $\langle T_i \rangle$ and thresholds N_{E_i} used for the burst selection.

substrate	N_{E_1}	$\langle T_1 \rangle$ [ms]	N_{E_2}	$\langle T_2 \rangle$ [ms]
none	>0.8	6.4	<0.3	8.7
Mg_2ATP	>0.7	5.9	<0.2	7.0
KADP	>0.8	7.0	<0.2	6.1
3-PG	>0.8	7.2	<0.2	6.1
3-PG*KADP	>0.8	7.3	<0.25	6.2

E.7 A Coarse-Grained Model of PGK Interdomain Dynamics

The interdomain dynamics of PGK was modeled with an elastic network [112] simulated under the random Multiparticle Collision Dynamics (rMPC) approach [31]. Here, the protein is described with Molecular Dynamics (MD) and the solvent with a heat bath under the random MPC approach [113][114] in a similar manner to Langevin Dynamics [115]. Moreover,

the normal modes analysis was performed with MMTK [37] under the same elastic network description. The analysis of the simulated trajectories was performed by mounting over the trajectory the rigid domains taken from the crystal (pdb:1qpg) in order to remove the contribution of the breathing modes. The domains were previously identified by DomainFinder [112]. Finally, the dyes were simulated with the accessible volume (AV) algorithm [73] in order to compare the measured physical quantities with the simulations. Further details about the simulations are reported in the following sections.

E.7.1 The Elastic Network Model of PGK

Under the elastic network approach [112], the protein is described by a network of atoms interconnected by springs up to a cut off distance r_c. In the present work, the network was built up with the C_α atoms only. Here, the potential energy, that gives rise to the force field used in the MD simulation to compute the trajectories, and employed to calculate the normal modes is:

$$U(\mathbf{r}_1, ..., \mathbf{r}_N) = \sum U_{ij}(\mathbf{r}_i - \mathbf{r}_j) \tag{E.21}$$

Where, the sum is over all the pairs of C_α atoms interacting through an harmonic potential:

$$U_{ij}(\mathbf{r}) = k\left(\mathbf{R}^0_{ij}\right)\left(|\mathbf{r}| - |\mathbf{R}^0_{ij}|\right)^2 \tag{E.22}$$

Where, $\mathbf{R}^0_{ij} = \mathbf{R}^0_i - \mathbf{R}^0_j$ is the distance vector between the atoms i and j of the reference configuration, that is assumed to be in a local mimimum, and $\mathbf{r} = \mathbf{r}_i - \mathbf{r}_j$ is a distance vector between the same pair of atoms. The previous equation shows that the character/geometry/direction of the motions depends exclusively on the topology of the network, which is defined by the protein structure through the pair interresidues contacts. On the other hand, the amplitude of the protein movements is encoded in the strength of the spring constants k. In the present work, an exponential decaying force constant between atoms pairs separated by a distance r lower than the cut off length ($|r| < r_C$) was used:

$$k(\mathbf{r}) = k_0 \cdot e^{-\frac{|\mathbf{r}|^2}{r_0^2}} \tag{E.23}$$

Where, r_0 is the persistence length of the interaction and k_0 an arbitrary constant. Here, the persistence length r_0 and the cut off length r_c were set to the values reported in Tab.E.15 [112]. Furthermore, the parameters k_0 was adjusted in order to match within $\sim 10\,\%$ the amplitude $\sigma_{CC} = 7.6\,\text{Å}$ of the fluctuations of the distance between the labeled C_α (S290 and Q135) which were measured by Haran et al [8]. Therefore, the model is independent from the experimental results presented in the current work. Finally, the reference configuration was derived from the

crystal structure of PGK from yeast (pdb:1qpg) by deleting all atoms except the C_α. Here, prior to the simulation, the quality of the two available crystal structures of PGK from yeast (3pgk and 1qpg) were checked by a Ramachandran plot and by a visual inspection. The older structure (3pgk) was discarded, because of the high number of residues in the outlier region of the Ramachandran plot ($\sim 21\,\%$). The parameters used for the simulation are reported in Tab.E.15.

Table E.15: Input parameters used to simulate PGK from yeast with an elastic network.

pdb	k_0	$[k_B T \cdot Å^{-2}]$	r_0 [Å]	r_c [Å]
1qpg		0.28	7	12

E.7.2 Random MPC Simulations

The trajectories of the C_α atoms were computed with the random MPC methods [31] using the potential energy described by Eq.E.21 and the parameters reported in Tab.E.15. Afterwards, the calculated trajectories were projected on the normalized vectors of the first three normal modes (hinge bending, propelled twist and rocking) previously calculated with MMTK [37]. For each trajectory, the configuration of the C_α-atoms was saved every 5000 integration steps. For each of these snapshot of the protein structure, the three domains (N-/C-terminus and hinge), previously identified by DomainFinder [112], were superimposed rigidly with a full atoms description, which comprise the substrates, by minimizing the RMSD between the C_α-atoms. Therefore, the breathing modes of the domains are removed. Finally, for each snapshot, the interesting physical quantities (R_g, r_{CC}, r_{PP}, $\langle r_{DA} \rangle_E$, $\langle r_{DA} \rangle$ and $\langle E \rangle$) were calculated (see Appx.E.7.4).

E.7.3 The Accessible Volume Description of the Dyes

The dyes contribution to the distance distributions were described with the accessible volume algorithm (AV) [73][33] in order to compare the model with the experimental results. The algorithm identifies the volume accessible by the dyes, represented as ellipsoids, under the constrains of the linkers and of an hard sphere potential. The accessible volumes were calculated for a set of conformations (see Appx.E.7.2) along the trajectories obtained from the random MPC simulations. The same parameters (Tab.E.16) used by Seidel et al [33] for dyes with almost the same chemical structure were used. The only difference resides on the lengths L_{link} of the linkers which were evaluated with VMD by taking the distance between the bottom of the linker and the center of the dye. A grid resolution of 0.95 Å was utilized to evaluate the AV.

Table E.16: Parameters used to calculate the accessible volumes. The dyes are represented as ellipsoids with three axial dimensions $R_{dye(i)}$. The linkers has a length L_{link} and a width w_{link} [33].

dye	$R_{dye(1)}$ [Å]	$R_{dye(2)}$ [Å]	$R_{dye(3)}$ [Å]	w_{link} [Å]	L_{link} [Å]
Al488	5	4.5	1.5	4.45	16.17
Al647	11	3	1.5	4.45	10.52

E.7.4 Calculation of the Physical Parameters

The interesting physical parameters (R_g, r_{CC}, r_{PP}, $\langle r_{DA}\rangle_E$, $\langle r_{DA}\rangle$ and $\langle E\rangle$) were calculated for each snapshot of the PGK structure along the trajectory simulated with the random MPC approach. The radius of gyration was computed employing only the C_α atoms positions [116]:

$$R_g^2 \simeq \frac{1}{N} \cdot \sum_i^N (\mathbf{r}_i - \mathbf{R}_{CM})^2 \tag{E.24}$$

Where, $\mathbf{r}_i$ is the position of the i-atom and $\mathbf{R}_{CM}$ is the center of mass position. The distances between the labeling positions r_{CC} (Q135 and S290) and the reactive phospates r_{PP} of the two substrates (3-PG and AMP-PNP) were calculated with a vectorial difference. The following equations were applied to calculate the mean donor-acceptor distance $\langle r_{DA}\rangle$, the FRET averaged distance $\langle r_{DA}\rangle_E$, and mean efficiency $\langle E\rangle$ [33][116]:

$$\langle r_{DA}\rangle = \langle |\mathbf{r}_{D(i)} - \mathbf{r}_{A(j)}| \rangle_{i,j} = \frac{1}{nm} \cdot \sum_i^n \sum_j^m |\mathbf{r}_{A(j)} - \mathbf{r}_{D(i)}| \tag{E.25}$$

$$\langle r_{DA}\rangle_E = R_0 \cdot \left[\frac{1}{\langle E\rangle} - 1 \right]^{\frac{1}{6}} \tag{E.26}$$

$$\langle E\rangle = \frac{1}{nm} \cdot \sum_i^n \sum_j^m \left[\frac{1}{1 + \left(\frac{|\mathbf{r}_{A(j)} - \mathbf{r}_{D(i)}|}{R_0} \right)^6} \right] \tag{E.27}$$

Here, all the position $\mathbf{r}$ within the accessible volume are considered to be equally probable.

Bibliography

[1] D. L. Nelson and M. M. Cox. *Lehninger Principles of Biochemistry - Third Edition*. Worth Publishers, 41 Madison Avenue, New York, 2000.

[2] T. N. Bryant, H. C. Watson, and P. L. Wendell. Structure of yeast phosphoglycerate kinase. *Nature*, 247:14–17, January 1974.

[3] R. D. Banks, C. C. F. Blake, P. R. Evans, R. Haser, D. W. Rice, G. W. Hardy, M. Merret, and A. W. Phillips. Sequence, structure and activity of phosphoglycerate kinase: A possible hinge-bending enzyme. *Nature*, 279:773–777, June 1979.

[4] T. M. McPhillips, B. T. Hsu, M. A. Sherman, M. T. Mas, and D. C. Rees. Structure of the R65Q mutant of yeast 3-phosphoglycerate kinase complexed with Mg-AMP-PNP and 3-phospho-D-glycerate. *Biochemistry*, 35:4118–4127, 1996.

[5] B. E. Bernstein, P. A. M. Michels, and W. G. J. Hol. Synergistic effects of substrate-induced conformational changes in phosphoglycerate kinase activation. *Nature*, 285:275–278, 1997.

[6] C. A. Pickover, D. B. McKay, D. M. Engelman, and T. A. Steitz. Subsrate binding closes the cleft between the domains of yeast phosphoglycerate kinase. *The Journal of Biological Chemistry*, 254:11323–11329, November 1979.

[7] A. Varga, B. Flachner, P. Konarev, E. Gráczer, J. Szabó, D. Svergun, P. Závodszky, and M. Vas. Substrate-induced double sided H-bond network as a means of domain closure in 3-phosphoglycerate kinase. *FEBS Letters*, 580:2698–2706, 2006.

[8] G. Haran, E. Haas, B. K. Szpikowska, and M. T. Mas. Domain motions in phosphoglycerate kinase: Determination of interdomain distance distribution by site-specific labeling and time-resolved fluorescence energy transfer. *PNAS*, 89:11764–11768, December 1992.

[9] Z. Palmai, L. Chaloin, C. Lionne, J. Fidy, D. Perahia, and E. Balog. Substrate binding modifies the hinge bending characteristics of human 3-phosphoglycerate kinase: A molecular dynamics study. *Proteins*, 77:319–329, 2009.

[10] E. Balog, M. Laberge, and J. Fidy. The influence of interdomain interactions on the intradomain motions in yeast phosphoglycerate kinase: A molecular dynamics study. *Biophysical Journal*, 92:1709–1716, March 2007.

[11] R. Inoue, R. Biehl, T. Rosenkranz, J. Fitter, M. Monkenbusch, A. Radulescu, B. Farago, and D. Richter. Large domain fluctuations on 50-ns timescale enable catalytic activity in phosphoglycerate kinase. *Biophysical Journal*, 99:2309–2317, 2010.

[12] N. Smolin, R. Biehl, G. R. Kneller, D. Richter, and J. C. Smith. Functional domain motions in proteins on the $\sim$ 1-100 ns timescale: Comparison of neutron spin-echo spectroscopy of phosphoglycerate kinase with molecular-dynamics simulation. *Biophysical Journal*, 102:1108–1117, March 2012.

[13] J. A. McCammon, B. R. Gelin, and M. Karplus. The hinge-bending mode in lysozyme. *Nature*, 262:325–326, July 1976.

[14] K. A. Dill and H. S. Chan. From Levinthal to pathways to funnels. *Nature Structural Biology*, 4:10–19, January 1997.

[15] S. Kumar, B. Ma, C. J. Tsai, N. Sinha, and R. Nussinov. Folding and binding cascades: Dynamic landscapes and population shifts. *Protein Sciences*, 9:10–19, 2000.

[16] D. D. Boehr, R. Nussinov, and P. E. Wright. The role of dynamic conformational ensembles in biomolecular recognition. *Nature Chemical Biology*, 5:789–796, December 2009.

[17] L. Meireles, M. Gur, A. Bakan, and I. Bahar. Pre-existing soft modes of motion uniquely defined by native contact topology facilitate ligand binding to proteins. *Protein Science*, 00:1–14, 2011.

[18] A. Bakan and I. Bahar. The intrinsic dynamics of enzymes plays a dominant role in determining the structural changes induced upon inhibitor binding. *PNAS*, 106:14349–14354, August 2009.

[19] I. Bahar, T. R. Lezon, A. Bakan, and I. H. Shirivastava. Normal mode analysis of biomolecular structures: Functional mechanisms of membrane proteins. *Chem. Rev.*, 110:1463–1497, 2010.

[20] K. Okazaki and S. Takada. Dynamic energy landscape view of coupled binding and protein comformational change: Induced-fit versus population-shift mechanisms. *PNAS*, 105:11182–11187, August 2008.

[21] O. Miyashita, J. N. Onuchic, and P. G. Wolynes. Nonlinear elasticity, proteinquakes, and the energy landscapes of functional transitions in proteins. *PNAS*, 100:12570–12575, October 2003.

[22] K. A. Henzler-Wildman, M. Lei, V. Thai, S. J. Kerns, M. Karplus, and D. Kern. A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature*, 262:913–916, July 2007.

[23] K. A. Henzler-Wildman, V. Thai, M. Lei, M. Ott, M. Wolf-Watz, T. Fenn, E. Pozharski, M. A. Wilson, G. A. Petsko, M. Karplus, C. G. Hübner, and D. Kern. Instrinsic motions along an enzymatic reaction trajectory. *Nature*, 450:838–844, December 2007.

[24] D. U. Ferreiro, J. A. Hegler, E. A. Komives, and P. G. Wolynes. On the role of frustration in the energy landscapes of allosteric proteins. *PNAS*, 108:3499–3503, March 2011.

[25] G. E. Schulz, C. W. Müller, and K. Diederichs. Induced-fit movements in adenylate kinases. *J. Mol. Biol.*, 213:627–630, 1990.

[26] H. J. Zhang, X. R. Sheng, X. M. Pan, and J. M. Zhou. Activation of adenylate kinase by denaturant is due to the increasing conformational flexibility at its active sites. *Biochemical and Biophysical Research Communcations*, 238:382–386, 1997.

[27] J. A. Hanson, K. Duderstadt, L. P. Watkins, S. Bhattacharyya, J. Brokaw, J. W. Chu, and H. Yang. Illuminating the machanistic roles of enzyme conformational dynamics. *PNAS*, 104:18055–18060, November 2007.

[28] K. Arora and C. L. Brooks III. Large-scale allosteric conformational transitions of adenylate kinase appear to involve a population-shift mechanism. *PNAS*, 104:18496–18501, November 2007.

[29] S. Nagarajan, D. Amir, A. Grupi, D. P. Goldenberg, A. P. Minton, and E. Haas. Modulation of functionally significant conformational equilibria in adenylate kinase by high concentrations of trimethylamine oxide attributed to volume exclusion. *Biophysical Journal*, 100:2991–2999, June 2011.

[30] J. Ådén, A. Verma, A. Schug, and M. Wolf-Watz. Modulation of a pre-existing conformational equilibrium tunes adenylate kinase activity. *JACS*, 134:16562–16570, October 2012.

[31] M. Ripoll, R. G. Winkler, and G. Gompper. Hydrodynamic screening of star polymers in shear flow. *Eur. Phys. J. E*, 23:349–354, 2007.

[32] Tobias Rosenkranz. *Time-resolved single molecule FRET studies on folding/unfolding transitions and on functional conformational changes of phosphoglycerate kinase*. PhD thesis, Heinrich Heine Universität Düsseldorf, 2011.

[33] S. Sindbert, S. Kalinin, H. Nguyen, A. Kienzler, L. Clima, W. Bannwarth, B. Appel, S. Müller, and C. A. M. Seidel. Accurate distance determination of nucleic acids via Förster resonance energy transfer: Implications of dye linker length and rigidity. *JACS*, 133:2463–2480, February 2011.

[34] E. A. Lemke. Site-specific labeling of proteins for single-molecule FRET measurements using genetically encoded ketone functionalities. *Bioconjugated Protocols: Strategies and Methods*, 257:3–15, 2011.

[35] Daryan Kempe. Time resolved fluorescence studies on proteins: Measurements and data analysis of anisotropy decays and single molecule FRET efficiencies. Master's thesis, Universität zu Köln, 2012.

[36] G. Allodi. FMINUIT - A binding to Minuit for Matlab, Octave and Scilab. `http://www.fis.unipr.it/~giuseppe.allodi/Fminuit/Fminuit_intro.html`, 1996-2010.

[37] K. Hinsen. *MMTK user's guide*. Centre de Biophysique Moleculaire, CNR Orleans, France.

[38] M. P. Lillo, J. M. Beechem, B. K. Szpikowska, M. A. Sherman, and M. T. Mas. Design and characterization of a multisite fluorescence energy-transfer system for protein folding studies: A steady-state and time-resolved study of yeast Phosphoglycerate kinase. *Biochemistry*, 36:11261–11272, 1997.

[39] V. Ratner, E. Kahana, M. Eichler, and E. Haas. A general strategy for site-specific double labeling of globular proteins for kinetic FRET studies. *Bioconjugated Chem.*, 13:1163–1170, 2002.

[40] E. Castiglioni. *CD spectropolarimetry: The instrumental approach*. E.C.S. European Chirality Services s.r.l., Cernusco Lambardone, Italy.

[41] S. M. Kelly and N. C. Price. The use of circular dichroism in the investigation of protein structure and function. *Current Protein and Peptide Science*, 1:349–384, 2000.

[42] Ch. Zander, J. Enderlein, and R. A. Keller. *Single Molecule Detection in Solution - Methods and Applications*. Wiley-VCH, Berlin, Germany, 2002.

[43] J. R. Fries, L. Brand, C. Eggeling, M. Köllner, and C. A. M. Seidel. Quantitative identification of different single molecules by selective time-resolved confocal fluorescence spectroscopy. *J. Phys. Chem. A*, 102:6601–6613, 1998.

[44] M. Antonik, S. Felekyan, A. Gaisuk, and C. A. M. Seidel. Separating structural heterogeneites from stochastic variations in fluorescence resonance energy transfer distributions via photon distribution analysis. *J. Phys. Chem. B*, 110:6970–6978, 2006.

[45] S. T. Hess and W. W. Webb. Focal volume optics and experimental artifacts in confocal fluorescence correlation spectroscopy. *Biophysical Journal*, 83:2300–2317, October 2002.

[46] S. Rüttinger, V. Buschmann, B. Krämer, R. Erdmann, R. MacDonald, and F. Koberling. Comparison and accuracy of methods to determine the confocal volume for quantitative fluorescence correlation spectroscopy. *Journal of Microscopy*, 232:343–352, May 2008.

[47] M. Wahl. Time-correlated single photon counting. *PicoQuant technical note*, 2009.

[48] M. T. Tyn and T. W. Gusek. Prediction of diffusion coefficients of proteins. *Biotechnology and Bioengineering*, 35:327–338, 1990.

[49] L. He and B. Niemeyer. A novel correlation for protein diffusion coefficients based on molecular weight and radius of gyration. *Biotechnol. Prog.*, 19:544–548, 2003.

[50] P. Kapusta. *Absolute diffusion coefficients: Compilation of reference data for FCS calibration*. PicoQuant GmbH, Berlin, Germany.

[51] Diaa Atta. *Time resolved single molecule fluorescence spectroscopy on surface tethered and freely diffusing proteins*. PhD thesis, Heinrich Heine Universität Düsseldorf, 2011.

[52] J. R. Lakowicz. *Principles of Fluorescence Spectroscopy - Second Edition*. Kluwert Academic/Plenum Publ., New York, USA, 1999.

[53] I. V. Gopich and A. Szabo. Theory of the energy transfer efficiency and fluorescence lifetime distribution in single-molecule FRET. *PNAS*, 109:7747–7752, May 2012.

[54] J. Wolberg. *Data Analysis Using the Method of Least Squares*. Springer, Berlin Heidelberg, Germany, 2006.

[55] L. Brand, C. Eggeling, C. Zander, K. H. Drexhage, and C. A .M Seidel. Single-molecule identification of Coumarin-120 by time-resolved fluorescence detection: Comparison of one- and two-photon excitation in solution. *J. Phys. Chem. A*, 101:4313–4321, 1997.

[56] J. Enderlein, P. M. Goodwin, A. Van Orden, W. P. Ambrose, R. Erdmann, and R. A. Keller. A maximum likelihood estimator to distinguish single molecules by their fluorescence decays. *Chemical Physical Letters*, 270:464–470, 1997.

[57] J. Tellinghuisen, P. M. Goodwin, W. P. Ambrose, J. C. Martin, and R. A. Keller. Analysis of fluorescence lifetime data for single Rhodamine molecules in flowing sample streams. *Anal. Chem.*, 66:64–72, 1994.

[58] E. Sisamakis, A. Valeri, S. Kalinin, P. J. Rothwell, and C. A. M. Seidel. Accurate single-molecule FRET studies using multiparameter fluorescence detection. *Methods in Enzymology*, 475:455–514, 2010.

[59] B. K. Müller, E. Zaychikov, C. Bräuchle, and D. C. Lamb. Pulsed interleaved excitation. *Biophysical Journal*, 89:3508–3522, November 2005.

[60] C. A. A. Vargas. moving_average v3.1. `http://www.mathworks.com/matlabcentral/fileexchange/12276-movingaverage-v3-1-mar-2008`, 2008.

[61] J. Enderlein, D. L. Robbins, W. P. Ambrose, P. M. Goodwin, and R. A. Keller. The statistics of single molecule detection: An overview. *Bioimaging*, 5:88–98, 1997.

[62] J. Widengren, V. Kudryavtsev, M. Antonik, S. Berger, M. Gerken, and C. A. M. Seidel. Single-molecule detection and identification of multiple species by multiparameter fluorescence detection. *Anal. Chem.*, 78:2039–2050, 2006.

[63] I. V. Gopich and A. Szabo. Single-molecule FRET with diffusion and conformational dynamics. *J. Phys. Chem. B*, 111:12925–12932, 2007.

[64] I. V. Gopich and A. Szabo. Theory of photon statistics in single-molecule Förster resonance energy transfer. *The Journal of Chemical Physics*, 122:014707, 2005.

[65] I. V. Gopich. Concentration effects in "single-molecule" spectroscopy. *J. Phys. Chem. B*, 112:6214–6220, 2008.

[66] T. Förster. Zwischenmolekulare Energiewanderung und Fluoreszenz. *Ann. Phys.*, 437:55–75, 1948.

[67] M. Cristóvão, E. Sisamakis, M. M. Hingorani, A. D. Marx, C. P. Jung, P. J. Rothwell, C. A. M. Seidel, and P. Friedhoff. Single-molecule multiparameter fluorescence spectroscopy reveals directional MutS binding to mismatched bases in DNA. *Nucleic Acids Research*, 2012.

[68] S. Kalinin, E. Sisamakis, S. W. Magennis, S. Felekyan, and C. A. M. Seidel. On the origin of broadening of single-molecule FRET efficiency distribution beyond shot noise limits. *J. Phys. Chem. B*, 114:6197–6206, 2010.

[69] I. V. Gopich and A. Szabo. FRET efficiency distribution of multistate single molecules. *J. Phys. Chem. B*, 114:15221–15226, 2010.

[70] S. Kalinin, S. Felekyan, A. Valeri, and C. A. M. Seidel. Characterizing multiple molecular states in single-molecule multiparameter fluorescence detection by probability distribution analysis. *J. Phys. Chem. B*, 112:8361–8374, 2008.

[71] Y. Santoso, J. P. Torella, and A. N. Kapanidis. Characterizing single-molecule FRET dynamics with probability distribution analysis. *ChemPhysChem*, 11:2209–2219, 2010.

[72] S. Kalinin, A. Valeri, M. Antonik, S. Felekyan, and C. A. M. Seidel. Detection of structural dynamics by FRET: A photon distribution and fluorescence lifetime analysis of systems with multiple states. *J. Phys. Chem. B*, 114:7983–7995, 2010.

[73] Simon Sindbert. *FRET restreined high-precision structural modeling of biomolecules*. PhD thesis, Heinrich Heine Universität Düsseldorf, 2012.

[74] F. James. *Statistical Methods in Experimental Physics - 2nd Edition*. World Scientific Publishing Co. Pte. Ltd., Singapore, 2006.

[75] J. N. Demas and G. A. Crosby. The measurement of photoluminescence quantum yields, a review. *The Journal Of Physical Chemistry*, 75:991–1024, April 1971.

[76] A guide to recording fluorescence quantum yields. Technical report, Horiba, Jobin Yvon Ltd., 2 Dalston Gardens, Stanmore, Middlesex, UK.

[77] R. Sens and K. H. Drexhage. Fluorescence quantum yield of oxazine and carbazine laser dyes. *Journal of Luminescence*, 24/25:709–712, 1981.

[78] G. Weber and F. W. Teale. Determination of the absolute quantum yield of fluorescent solutions. *Trans. Faraday Soc.*, 53:646–655, 1957.

[79] H. Boroudjerdi, A. Naji, and R. R. Netz. *DNA and Chromatin*, chapter C9. 42nd - IFF Springschool, Forschugszentrum Jülich.

[80] D. Tobi and I. Bahar. Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state. *PNAS*, 102:18908–18913, December 2005.

[81] E. Freire. Statistical thermodynamics linkage between conformational and binding equilibria. *Advances in Protein Chemistry*, 51:255–279, 1998.

[82] L. Zerrad, A. Merli, G. F. Schröder, A. Varga, E. Gráczer, P. Pernot, A. Round, M. Vas, and M. W. Bowler. A spring-loaded release mechanism regulates domain movement and catalysis in phosphoglycerate kinase. *Journal of Biological Chemistry*, 286:14040–14048, 2011.

[83] J. Enderlein, I. Gregor, D. Patra, T. Dertinger, and U. B. Kaupp. Performance of fluorescence correlation spectroscopy for measuring diffusion and concentration. *ChemPhysChem*, 6:2324–2336, 2005.

[84] P. R. Selvin and T. Ha. *Single-Molecule Techniques - A laboratory manual*. Cold Spring Harbor laboratory press, Cold Spring Harbor, New York, 2008.

[85] A. Dhar, A. Samiotakis, S. Ebbinhaus, L. Nienhaus, D. Homouz, M. Gruebele, and M. S. Cheung. Structure, function, and folding of phosphoglycerate kinase are strongly perturbed by macromolecular crowding. *PNAS*, 107:17586–17591, October 2010.

[86] T. Rosenkranz, R. Schlesinger, M. Gabba, and J. Fitter. Native and unfolded states of phosphoglycerate kinase studied by single-molecule FRET. *ChemPhysChem*, 12:704–710, 2011.

[87] C. J. Tsai, B. Ma, and R. Nussinov. Folding and binding cascades: shifts in energy landscapes. *PNAS*, 96:9970–9972, August 1999.

[88] T. Plass, S. Milles, C. Koehler, C. Schultz, and E. A. Lemke. Genetically encoded copper-free click chemistry. *Angew. Chem. Int. Ed.*, 50:3878–3881, 2011.

[89] T. Bücher. Phosphoglycerate kinase from Bruwe's yeast. *Biochim. et Biophys. Acta 1*, 292:415–422, 1947.

[90] S. F. Velick and J. E. Hayes Jr. Phosphate binding and the glyceraldehyde-3-phosphate dehydrogenase reaction. *J. Biol. Chem.*, 203:545–562, 1953.

[91] I. Luque, S. A. Leavitt, and E. Freire. The linkage between protein folding and functional cooperativity: two sides of the same coin? *Annu. Rev. Biophys. Biomol. Struct.*, 31:235–256, 2002.

[92] J. Hohlbein, L. Aigrain, T. D. Craggs, O. Bermek, O. Potapova, P. Shoolizadeh, N. D. F. Grindley, C. M. Joyce, and A. N. Kapanidis. Conformational landscapes of DNA polymerase and mutator derivatives establish fidelity checkpoints for nucleotide insertion. *Nature Communications*, 4:2131, July 2013.

[93] S. R. Akabayov, B. Akabayov, C. C. Richardson, and G. Wagner. Molecular crowding enhanced ATPase activity of the RNA helicase eIF4A correlates with compaction of its quaternary structure and association with eIF4G. *JACS*, 135:10040–10047, 2012.

[94] R. Orij, J. Postmus, A. T. Beek, S. Brul, and G. J. Smits. In vivo measurements of cytosolic and mitochondrial pH using a pH sensitive GFP derivative Saccharomyces cerevisiae reveals a relation between intraellular pH and growth. *Microbiology*, 155:268–278, 2009.

[95] A. Ansari, C. M. Jones, E. R. Henry, J. Hofrichter, and W. A. Eaton. The role of solvent viscosity in the dynamics of protein conformational changes. *Science*, 256:1796–1798, June 1992.

[96] I. V. Gopich and A. Szabo. Decoding the pattern of photon colors in single-molecule FRET. *J. Phys. Chem. B*, 113:10965–10973, 2009.

[97] H. S. Chung, I. V. Gopich, K. McHale, T. Cellmer, J. M. Louis, and W. A. Eaton. Extracting rate coefficients from single-molecule photon trajectories and FRET efficiency histograms for a fast-folding protein. *J. Phys. Chem. A*, 115(16):3642–3656, 2011.

[98] J. P. Torella, S. J. Holden, Y. Santoso, J. Hohlbein, and A. N. Kapanidis. Identifying molecular dynamics in single-molecule FRET experiments with burst variance analysis. *Biophysical Journal*, 100:1568–1577, 2011.

[99] M. Margittai, J. Widengren, E. Schweinberger, G. F. Schröder, S. Felekyan, E. Haustein, M. König, D. Fasshauer, H. Grubmüller, R. Jahan, and C. A. M. Seidel. Single-molecule fluorescence resonance energy transfer reveals a dynamic equilibrium between closed and open conformations of syntaxin 1. *PNAS*, 100:15516–15521, 2003.

[100] M. Bömer, M. Wahl, H. J. Rahn, R. Erdmann, and J. Enderlein. Time-resolved fluorescence correlation spectroscopy. *Chemical Physics Letters*, 353:439–445, 2002.

[101] S. Felekyan, S. Kalinin, H. Sanabria, A. Valeri, and C. A. M. Seidel. Filtered FCS: species auto- and cross-correlation functions highlight binding and dynamics in biomolecules. *ChemPhysChem*, 13:1036–1053, 2012.

[102] T. Ha and P. Tinnefeld. Photophysics of fluorescent probes for single-molecule biophysics and super-resolution imaging. *Annu. Rev. Phys. Chem*, 63:597–617, 2012.

[103] J. Vogelsang, R. Kasper, C. Steinhauer, B. Person, M. Heilemann, M. Sauer, and P. Tinnefeld. A reducing and oxidizing system minimizes photobleaching and blinking of fluorescent dyes. *Angew. Chem. Int. Ed.*, 47:5465–5469, 2008.

[104] L. A. Campos, J. Liu, X. Wang, R. Ramanathan, D. S. English, and V. Muñoz. A photoprotection strategy for microsecond-resolution single-molecule fluorescence spectroscopy. *Nature Methods*, 8:143–146, 2011.

[105] J. Shaffer, A. Volmer, C. Eggeling, V. Subramaniam, G. Striker, and C. A. M. Seidel. Identification of single-molecule in aqueous solution by time-resolved fluorescence anisotropy. *J. Phys. Chem. A*, 103:331–336, January 1999.

[106] M. Koshioka, K. Sasaki, and H. Masuhara. Time-dependent fluorescence depolarization analysis in three-dimensional microspectroscopy. *Appl. Spect.*, 49:224–228, 1995.

[107] R. A. Velapoldi and H. H. Tønnesen. Corrected emission spectra and quantum yields for a series of fluorescence compounds in the visible spectral region. *Journal of Fluorescence*, 14:465–472, July 2004.

[108] S. Kalinin, T. Peulen, S. Sinbert, P. J. Rothwell, S. Berger, T. Restle, R. S. Goody, H. Gohlke, and C. A. M. Seidel. A toolkit and benchmark study for FRET-restrained high-precision structural modeling. *Nature Methods*, 9:1218–1225, 2012.

[109] G. Hummer. From transition paths to transition states and rate coefficients. *Journal of Chemical Physics*, 120:516–523, January 2004.

[110] P. Hänggi and P. Talkner. Reaction-rate theory: fifty years after Kramers. *Reviews of Modern Physics*, 62:252–332, April 1990.

[111] W. Kügel, A. Muschielok, and J. Michaelis. Bayesian-inference-based fluorescence correlation spectroscopy and single-molecule burst analysis reveal the influence of dye selection on DNA hairpin dynamics. *ChemPhysChem*, 13:1013–1022, 2012.

[112] K. Hinsen. Analysis of domain motions by approximate normal mode calculations. *Proteins*, 33(3):417–429, 1998.

[113] M. Ripoll. *Molecular Dynamics*, chapter A8. 42nd - IFF Springschool, Forschugszentrum Jülich.

[114] M. Ripoll. *Mesoscale Hydrodynamics*, chapter A10. 42nd - IFF Springschool, Forschugszentrum Jülich.

[115] G. Nägele. *Brownian Dynamics Simulations*, chapter B4.1-B5.3. 37th - IFF Springschool, Forschugszentrum Jülich.

[116] M. Y. Lobanov, N. S. Bogatyreva, and O. V. Lazitskaya. Radius of gyration as an indicator or protein compactness. *Molecular Biology*, 42:623–628, 2008.

[117] Émile Zola. *The Belly of Paris*. Oxford World's Classics, Oxford University Press Inc., New York, 2007.

[118] Émile Zola. *Il Ventre di Parigi*. Garzanti, Garzanti Libri S.r.l., Milano, 2007.

Summary

The protocols to perform reliable and reproducible single-molecule FRET (smFRET) experiments with a confocal microscope on diffusing molecules were established. This method was applied to investigate the interdomain functional dynamics of phosphoglycerate kinase (PGK) from yeast. In fact, this enzyme catalyzes an important step of the glycolytic pathway, where a reversible phosphor transfer reaction takes place between two substrates bound to the two different domains of PGK. Therefore, smFRET experiments were performed on freely diffusing PGK labeled on the two structural subunits, in order to characterize the timescale of the fluctuations and the distribution of states of the domains. Furthermore, the redistribution of the conformational substates upon ligand binding was investigated. Finally, an elastic network model of the protein was simulated under the random Multiparticle Collision Dynamics approach, including the dye linkers contribution with an accessible volume algorithm. The model excellently describes the smFRET experiments. Moreover, the functional relevance of the interdomain motions and their geometry were identified.

The results highlight the intrinsically dynamic and flexible nature of PGK, with dynamics spanning from nanoseconds to milliseconds. A partially closed and compact conformation of the domains was disclosed. Large amplitude Brownian fluctuations start from this partially closed state in the direction towards the catalytically competent conformation. These functionally relevant fluctuations are faster than milliseconds, most likely on the nanoseconds timescale, and are encoded in the structural topology of PGK. A set of open and extended conformations of the domains was also observed. These conformers are static on the milliseconds timescales, and their functional relevance is still not known. The open and the partially closed states interconvert on the milliseconds timescale. Ligand binding reshapes the states distribution populating the partially closed state which better allows to reach the catalytically competent conformation. In addition, a conformational rearrangement within the closed state increases the inter-dye distance by means of a propelled twist like motion.

These results have an impact on the present view on the PGK catalytic behavior, complementing and supporting previous works. In fact, the geometry of the interdomain motions, together with the possible functional relevance of these motions, is assessed. Furthermore, the results of this thesis support the conformational selection model of substrate binding, and the emerging

paradigm which states that *'structure-encodes-dynamics'* and *'dynamics-encodes-function'*. In fact, the ligands favour the conformation of the domains most suitable for function, which is the closed one, and the functional motions of PGK are encoded in the 3D structure of the ligand-free protein.

Zusammenfassung

Ziel dieser Arbeit war es, mittels konfokaler Fluoreszenz Mikroskopie verlässliche und reproduzierbare Einzelmolekül-FRET Untersuchungen an frei diffundierenden Molekülen zu etablieren und anzuwenden, um funktionell relevante Bewegungen der beiden Domänen der Phosphoglyceratkinase (PGK)aus der Hefe Saccharomyces cerevisiae zu untersuchen. Dieses Enzym katalysiert einen entscheidenden Schritt der Glykolyse, nämlich den reversiblen Phosphattransfer zwischen den zwei Substraten der PGK, die jeweils an eine der beiden Domänen binden. Um den Zeitrahmen der Fluktuationen und die Verteilung der Zustände der PGK-Domänen zu charakterisieren, wurden Einzelmolekül-FRET Experimente an frei diffundierenden Molekülen durchgeführt. Dazu wurde das Protein an beiden strukturellen Untereinheiten fluoreszenzmarkiert. Zusätzlich wurde die Neuverteilung der Konformationen während der Ligandenbindung untersucht. Mit Hilfe des random Multiparticle Collision Dynamics Ansatz wurde schließlich ein elastisches Netzwerkmodell des Proteins simuliert, das einen accessible volume-Algorithmus zur Berechnung des Farbstoff-Linker-Beitrags enthält. Das Modell beschreibt die Einzelmoleküluntersuchungen äußerst genau und dient darüber hinaus dazu, die funktionelle Relevanz der Domänenbewegungen und deren Geometrie aufzuklären. Die Daten spiegeln die flexible Struktur der PGK und ihre intrinsische Dynamik auf einer Zeitskala von Nano- bis Millisekunden wider. Sie zeigen eine partiell geschlossene und kompakte Konformation der Domänen. BROWN'sche Bewegungen großer Amplitude sind von diesem partiell geschlossenen Zustand bishin zur katalytisch aktiven Konformation zu beobachten. Diese funktionell relevanten Fluktuationen sind schneller als Millisekunden, sehr wahrscheinlich im Nanosekundenbereich, und liegen in der strukturellen Topologie der PGK begründet. Verschiedene offene und ausgedehnte Konformationen der Domänen konnten ebenfalls beobachtet werden. Diese Konformere sind für Millisekunden stabil, jedoch ist ihre funktionelle Relevanz bisher unbekannt. Die Änderung der Konformation vom offenen zum partiell geschlossenen Zustand und andersherum findet in wenigen Millisekunden statt. Ligandenbindung verschiebt die Verteilung der Konformationen in Richtung des partiell geschlossenen Zustands, der die Bildung der katalytisch aktiven Konformation begünstigt. Zusätzlich wird die Distanz zwischen beiden Farbstoffen durch eine propeller-ähnliche Drehbewegung während der Konformationsänderung innerhalb des geschlossenen Zustands der PGK erhöht. Die Ergebnisse dieser Arbeit

ergänzen und unterstützen frühere Erkenntnisse über das katalytische Verhalten der PGK. Die Geometrie der Domänenbewegungen und deren funktionelle Relevanz konnten gezeigt werden. Außerdem bekräftigen die Ergebnisse dieser Arbeit das Konformations-Selektions-Modell der Substratbindung und das aufkommende Paradigma, welches besagt, dass Struktur Dynamik und Dynamik Funktion kodiert. Und tatsächlich bevorzugen die Liganden die Konformation, die am ehesten funktionell ist und die funktionellen Bewegungen der PGK Domänen sind in der 3D Struktur des liganden-freien Proteins kodiert.

Acknowledgments

Here, I would like to thank Prof. Jörg Fitter, Prof. Georg Büldt, the BioSoft Graduate School in Biophysics and Soft Matter, and all my group... It is because of them if I could to do what I like in the best conditions. Then... I hope everyone can find out her/his own 'vegetable inside' which I would like to thank...

"It was like an ocean spreading between the two groups of markets from the Pointe Saint-Eustache to the Rue des Halles. In the two open spaces at either end the flood of greenery rose even higher, submerging the footpaths. Dawn came slowly, a soft grey that spread light watercolour tints everywhere. The piles of greenery were like waves, a river of green flowing along the roadway like an autumn torrent; and they assumed delicate, shadowy hues-pale violet, milky pink, and greenish yellow, all the soft, light hues that turn the sky into a canopy of shot silk as the sun rises. By degrees, as the fires of dawn rose higher and higher at the far end of the Rue Rambuteau, the mass of vegetables grew brighter and brighter, emerging more and more clearly from the bluish shadows on the ground. Lettuces, endives, chicory, open and with rich soil still clinging to their roots, exposed their swelling hearts; bunches of spinach, sorrel, and artichokes, piles of peas and beans, mounds of cos lettuces, tied up with straw, sounded every note in the scale of greens, from the lacquered green of the pods to the coarse green of the leaves; a continuous scale of rising and falling notes that died away in the mixed tones of the tufts of celery and the bundles of leeks. But the highest notes, at the very top of the scale, came from the bright carrots and snowy turnips, scattered in tremendous quantities throughout the markets, which they lit up with their medley of colours. At the intersection in the Rue des Halles, mountains of cabbages were piled up; there were enormous white ones, as hard as cannon balls, curly ones with big leaves that made them look like bronze bowls, and red ones which the dawn seemed to transform into magnificent flowers with the hue of wine-dregs, splashed with crimson and dark purple. On the other side of the markets, at the intersection near Saint-Eustache, the opening to the Rue Rambuteau was blocked by a barricade of orange pumpkins in two rows, sprawling at their ease and swelling out their bellies. Here and there gleamed the varnished golden-brown of a basket of onions, the blood-red of a heap of tomatoes, the soft yellow of a display of cucumbers, and the deep

mauve of aubergines; while large black radishes, laid down in funereal carpets, formed dark patches in the brilliance of the early morning." [117]

Ringraziamenti

Spero che ognuno possa riconoscere la propria 'verdura dentro' che vorrei ringraziare ...

"Era un mare. Un mare che si stendeva dall'angolo di Saint-Eustache fino alla strada dei mercati, tra due gruppi di padiglioni; e alla fine dei due incroci, l'onda si faceva piú alta e tutto il suolo era sommerso dalle verdure. Il giorno procedeva lento, soffuso di una luce grigia dolcissima, che stingeva i colori stemperando le tinte in una sfumatura d'acquarello. Quei cumuli accavallati come onde, quel fiume di verdura che pareva scorrere nel letto della strada come le piogge d'autunno quando si rovesciano improvvise, si vestivano di ombre leggere e perlate, di violetti teneri, di rose intinte nel latte, di verdi sciolti nel giallo, di tutti quei pallori che fanno del cielo una seta cangiante al levarsi del sole; e man mano che l'incendio del mattino divampava con raggi di fiamma, in fondo a rue Rambuteau, gli erbaggi si ravvivavano via via, emergendo da quella immensa e fredda nube azzurrina che lambiva ancora la terra. Le insalate, le lattughe, le indivie, i radicchi con le foglie dischiuse ed umide di terriccio, mostravano il loro cuore sgargiante; i mazzi di spinaci, di acetosella, di carciofi, i mucchi di fagioli, piselli, le torri di lattuga romana, legate con un fuscello, cantavano tutta la scala del verde, dalla lacca dei baccelli su fino al verdone delle foglie, gamma che andava estinguendosi nelle sfumature del gambo, dei sedani e dei porri. Ma le note piú acute, piú squillanti, erano pur sempre date dalle macchie accese della carota e da quelle pure della rapa; disseminate in quantitá enorme per tutto il mercato, mettevano allegria con il contrasto dei loro due colori. All'incrocio della strada dei mercati, c'erano montagne di cavoli: gli enormi cavoli bianchi chiusi e duri come palle di metallo, cavoli ricciuti dalle larghe foglie, simili a tazze di bronzo, cavoli viola che l'aurora trasformava in fiori stupendi, color feccia di vino con lividure di carminio e porpora scuro. Dalla parte opposta, all'incrocio di Saint-Eustache, l'inizio di rue Rambuteau era chiuso da una barricata di zucche, in due file, che facevano mostra dei loro larghi ventri. E la vernice rosso cangiante di un cesto di cipolle, il rosso sanguigno di un mucchio di pomodori, il giallo sbiadito di una partita di citrioli, il violetto cupo di un grappolo di melanzane, luccicavano e si accendevano qua e lá, mentre i grossi ramolacci, neri, appesi come nastri da lutto gettavano qualche squarcio scuro in mezzo a quel tripudio di colori e di vita." [118]

Band / Volume 67
Adsorption of (hetero-)organic Phthalocyanine and PTCDA thin films on Ag(111)
C. Kleimann (2013), iv, 149 pp
ISBN: 978-3-89336-888-4

Band / Volume 68
High-energy high-momentum surface spin waves of ultrathin epitaxial 3d transition metal films
R. Jayaraman (2013), ix, 135 pp
ISBN: 978-3-89336-890-7

Band / Volume 69
Magnetic Oxide Heterostructures: EuO on Cubic Oxides and on Silicon
C. Caspers (2013), xiii, 153 pp
ISBN: 978-3-89336-891-4

Band / Volume 70
Femtosecond Spin Dynamics in Magnetic Multilayers Employing High Harmonics of Laser Radiation
D. Rudolf (2013), vi, 121 pp
ISBN: 978-3-89336-894-5

Band / Volume 71
Development and application of a massively parallel KKR Green function method for large scale systems
A. R. Thieß (2013), ii, 173 pp
ISBN: 978-3-89336-906-5

Band / Volume 72
Conformational Dynamics of Calmodulin and Ribosome-Nascent Chain Complexes Studied by Time-Resolved Fluorescence Anisotropy
P. Lamprou (2013), 182 pp
ISBN: 978-3-89336-907-2

Band / Volume 73
Stereoselective synthesis of vicinal diols with enzymatic cascade reactions
J. Kulig (2013), XV, 177 pp
ISBN: 978-3-89336-911-9

Band / Volume 74
Computing Solids
Models, ab-initio methods and supercomputing
Lecture Notes of the 45th IFF Spring School 2014
March 10 – 21, 2014 Jülich, Germany
ed. by S. Blügel, N. Helbig, V. Meden, D. Wortmann (2014), ca. 1000 pp
ISBN: 978-3-89336-912-6

Band / Volume 75
Aberrationskorrigierte Photoemissionsmikroskopie an magnetischen Systemen: Von statischer Charakterisierung zu zeitaufgelöster Abbildung
F. Nickel (2013), x, 99 pp
ISBN: 978-3-89336-913-3

Band / Volume 76
Temperature-Induced Metamagnetic Transition and Domain Structures of Single-Crystalline FeRh Thin Films on MgO(100)
X. Zhou (2013), xi, 104 pp
ISBN: 978-3-89336-919-5

Band / Volume 77
Interplay between Magnetism and Superconductivity in Iron Based High Temperature Superconductors
S. Price (2013), 196 pp
ISBN: 978-3-89336-921-8

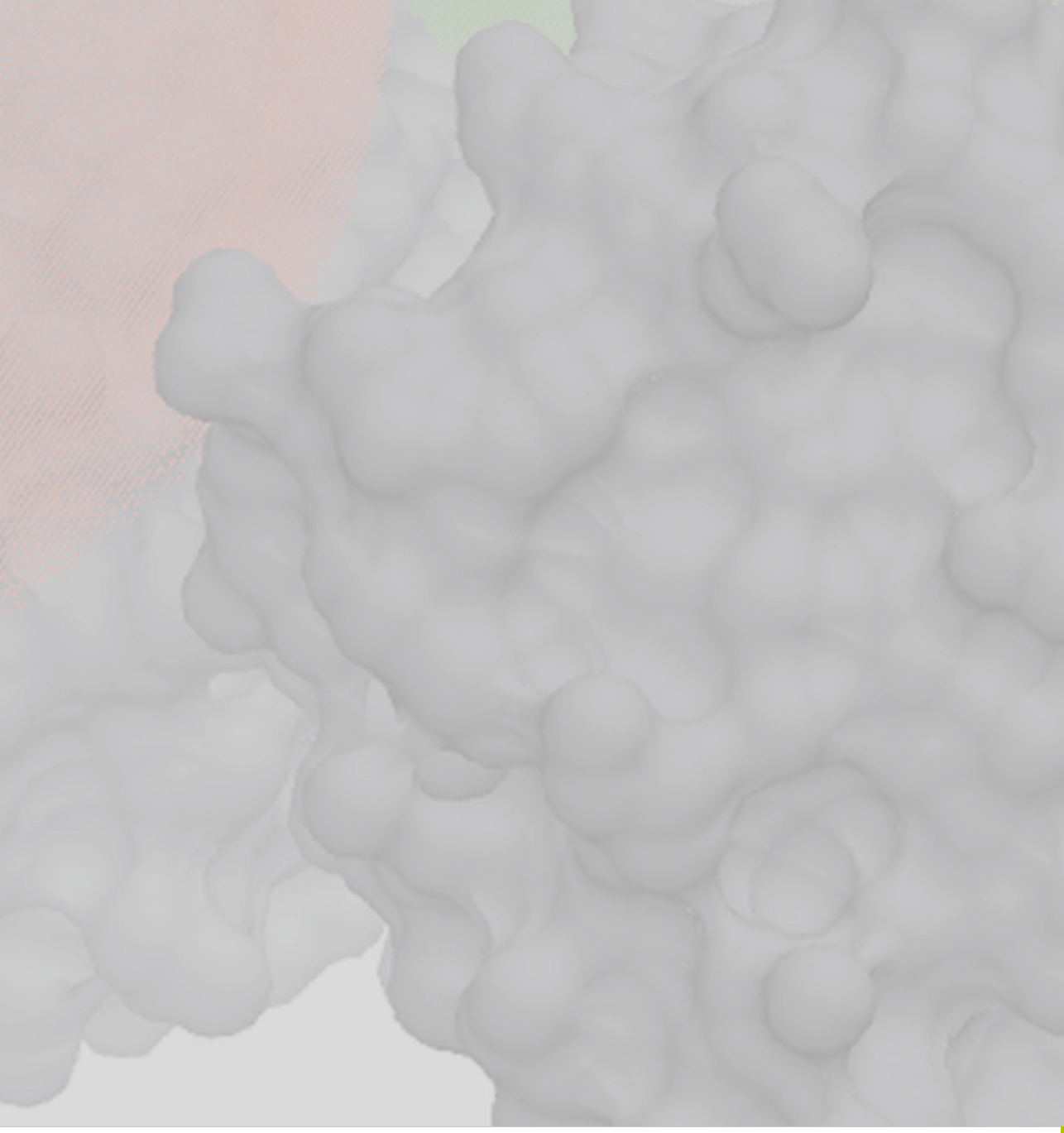
Band / Volume 78
Magnetoresistance and transport in carbon nanotube-based devices
C. Morgan (2013), viii, 131 pp
ISBN: 978-3-89336-926-3

Band / Volume 79
Development of a relativistic full-potential first-principles multiple scattering Green function method applied to complex magnetic textures of nano structures at surfaces
D. Bauer (2014), 193 pp
ISBN: 978-3-89336-934-8

Band / Volume 80
Identifizierung von artifiziellen Liganden eines in Nanodiscs inkorporierten integralen Membranproteins
M. Pavlidou (2014), 106 pp
ISBN: 978-3-89336-942-3

Band / Volume 81
Interdomain Functional Dynamics of Phosphoglycerate Kinase Studied by Single-Molecule FRET
M. Gabba (2014), v, 179 pp
ISBN: 978-3-89336-943-0

Weitere ***Schriften des Verlags im Forschungszentrum Jülich*** unter
http://wwwzb1.fz-juelich.de/verlagextern1/index.asp



Schlüsseltechnologien / Key Technologies
Band / Volume 81
ISBN 978-3-89336-943-0

JÜLICH
FORSCHUNGSZENTRUM